

## eAppendix

This package contains SAS programs that use inverse probability weighting approach to

1. estimate adherence-adjusted hazard ratios and their 95% confidence intervals (CIs).
2. estimate adherence-adjusted absolute risks and risk differences and their 95% CIs.
3. create adherence-adjusted, standardized outcome-free survival curves.

The package contains four SAS files:

### 1. Program 1 Data management

- Read in and manage datasets from the Women's Health Initiative Estrogen-plus-Progestin Trial, define variables that will be used in the analysis
- Create a SAS dataset (`ipwdr_bc.sas7bdat`) that will be used in later programs

### 2. Program 2 Weight and hazard ratio estimation

- Estimate stabilized weights
- Perform weighted analysis to estimate adherence-adjusted average hazard ratios and 95% CIs using bootstrapping

### 3. Program 3 Risk difference estimation

- Perform weighted analysis to estimate adherence-adjusted risk differences and 95% CIs using bootstrapping

### 4. Program 4 Survival curves

Create adherence-adjusted, standardized breast cancer-free survival curves

In this example:

The exposure of interest is postmenopausal estrogen-plus-progestin therapy.

The outcome of interest is invasive breast cancer.

The datasets from the Women's Health Initiative Estrogen-plus-Progestin Trial can be obtained through application from the National Heart, Lung, and Blood Institute at its Biologic Specimen and Data Repository Information Coordinating Center (<https://biolincc.nhlbi.nih.gov/home/>). If you have any question, comment, or discover an error, please contact Sengwee Darren Toh at [darrentoh@post.harvard.edu](mailto:darrentoh@post.harvard.edu). For the most updated versions of the SAS programs, please visit [www.hsph.harvard.edu/causal](http://www.hsph.harvard.edu/causal).

```
*****;
*
*                                     PROGRAM 1
*                                     *
*                                     DATA MANAGEMENT
*                                     *
* The raw datasets used in this Program can be obtained through application from the
* National Heart, Lung, and Blood Institute (NHLBI) at its Biologic Specimen and Data
* Repository Information Coordinating Center (https://biolincc.nhlbi.nih.gov/home/).
*                                     *
*****;
```

```
%MACRO data_management;
```

```
/*-----
  READ BASELINE DATA
  -----*/
```

```
/* demographic and study membership */
```

```
DATA dem_ep_base_pub;
INFILE "&file_location.dem_ep_base_pub.dat" LRECL=28;
INPUT id r dmflag cadflag age ager region ethnic lang educ1 incomel;
DROP lang;
RUN;
```

```
/* eligibility screening */
```

```
DATA f2_ep_base_pub;
INFILE "&file_location.f2_ep_base_pub.dat" LRECL=149;
INPUT id f2days area3y othstdy exstdy brca_f2 colon_f2 colon10y endo_f2 endo10y skinb
melan_f2 melan10y othcal10y race hearstdy horm hormnw horm3m osteobkb hormbk hystb
hyst3m hystageb mense1st mealout maldiet lfdietf2 diab diabage diabcoma dbdietf2
insulin insulinw diabpill diabetesb dvtb dvt6m dvtacclm peb pe6m peacc1m strokeb
stroke6m tiab tia6m mib miage mi6m scanemia chf_f2 liverdisb bldprobb l15lbs6m dialysis
othchronb hardstdy comecc helpcc intdm availdm inthrt availhrt talkdoc hrtinldr helpfill
age hrtb agehyst diabtrtb;
```

```
KEEP id hystb hystageb skinb mense1st dvtb peb strokeb
tiab mib liverdisb bldprobb othchronb inthrt availhrt talkdoc hrtinldr diabtrtb;
RUN;
```

```
/* personal information */
```

```
DATA f20_ep_base_pub;
INFILE "&file_location.f20_ep_base_pub.dat" LRECL=137;
INPUT id f20days educ notwrk retired homemkr employed disabled othwrk jobhmmkr jobmangr jobtech
jobserv joblabor joboth marital peduc pnotwrk pretired phomemkr pemploy pdisable pothwrk
pmainjob incomel careprov lstvisdy mammob lstmamy papsmearb lstpapy abnpap3y cervdys
endoasp lstaspdy hmoins lstrvin medicare medicaid mltryins noins payoth usserve
vamedctr mainjob nomam2yr nopap3yr timelast timelsts anyins;
KEEP id marital careprov mammob papsmearb abnpap3y cervdys endoasp nomam2yr nopap3yr timelsts anyins;
RUN;
```

```
/* medical history */
```

```
DATA f30_ep_base_pub;
INFILE "&file_location.f30_ep_base_pub.dat" LRECL=207;
INPUT id f30days hosp2y glaucomb cataractb hicholb asthmab emphysemb kidneystb hibldcab
stomulcrb diverticb colitisb lupusb pancreatb osteopb hiprepb othjrepb intestrm migraineb
alzheimb msb parkinsb alsb nacond cvdb arrestb chfb cathb cabgb ptcab carotidb afb aneurysmb
nacvd arthritb rheumat gallbsb gallbsnwb gallstrmb gallblrmb thyroidb goiter goiternv nodule
nodulenw ovrtthy ovrtthyw undthy undthyw hypt hyptage hyptpill hyptpiln anginab anginapb
padb padangr padangp padsurg colnsncpy colnsctdb pcolonrmb hemocculb hemocctdb canb brcab
brca55 colonb coloca55 thyrca thyrca55 cervcab skinca meln_f30 bladca otherca numfallsb
faintedb bkboneb bkhip bkhip55 bkback bkback55 bkuarm bkuarm55 bkfarm bkfarm55 bkhand
bkhand55 bkleg bkleg55 bkfoot bkfoot55 bkothb bkothb55 htnttrtb hip55 fract55 revascb;
KEEP id hosp2y glaucomb cataractb hicholb asthmab emphysemb kidneystb hibldcab stomulcrb diverticb
colitisb lupusb pancreatb osteopb hiprepb othjrepb migraineb alzheimb msb parkinsb alsb cvdb
arrestb chfb cathb cabgb ptcab carotidb afb aneurysmb arthritb gallbsb gallbsnwb gallstrmb
gallblrmb thyroidb anginab anginapb padb colnsncpy colnsctdb pcolonrmb hemocculb hemocctdb
canb brcab colonb cervcab numfallsb faintedb bkboneb htnttrtb revascb;
RUN;
```

```
/* reproductive history */
```

```
DATA f31_ep_base_pub;
INFILE "&file_location.f31_ep_base_pub.dat" LRECL=133;
INPUT id f31days menarcheb mensreg mensrega menopsea menswoly menswod anymensa menpsymb menpsyaf
menpsyal preg pregnum preg6m preg6mn preg6maf preg6mal brthlivn brthstln miscaryn ectpreg
nocnceiv nocncvdr nocncvr nocncvhr nocncvut nocncven nocncvot nocncvpt nocncvdk brstfeed
brstfdn brstfdaf brstfdal brstfdm oophb oophab tubtied tubtieda nedlaspb nedlaspa brstbiopb
brstbion brstaug brstauga brstaugs brstimp brstopot brstprem brstrem brstremo gravid parity
fulltrmr numliver agefbir boophb brstfdmo brstdisb;
KEEP id menarcheb menopsea anymensa menpsymb menpsyaf preg brstfeed brstfdn oophb oophab nedlaspb
brstbiopb gravid parity agefbir boophb brstdisb;
RUN;
```

```
/* family history */
```

```
DATA f32_ep_base_pub;
INFILE "&file_location.f32_ep_base_pub.dat" LRECL=173;
INPUT id f32days sister sisnum brother bronum daughter daunum son sonnum momalive momdieda momage
dadalive daddieda dadage diabrel diabreln mirel mimom midad misis1 misis2 misis3 mibro1 mibro2
mibro3 midaul midau2 mison1 mison2 strkrel strkreln cancrel cancrelncancrel brcas1 brcas12
brcas13 brcadaul brcadau2 brcadau3 brcagmam brcagmap colofrel colomom colosis1 colosis2 colosis3
colodaul colodau2 cervrel cervreln endorel endoreln ovarrel ovarreln cancmrel colomrel colodad
```

```

colobrol colobro2 colobro3 coloson1 coloson2 prosrel prosreln bkbonmom bkhipmom bkbckmom bkuarmom
bklarmom bktothmom bkbondad bkhipdad bkbckdad bkwardad bklardad bktothdad bkbonrel colorel;
KEEP id diabrel mirel mimom midad strkrel cancfrel brcafrel cervrel endorel ovarrel cancmrel
bkbonrel colorel;
RUN;

/* personal habits */
DATA f34_ep_base_pub;
INFILE "&file_location.f34_ep_base_pub.dat" LRECL=209;
INPUT id f34days smokevr smokage smoknow qsmokage qsmokhp cigsdaysb smokyrbsb smokwgt coffeeb cupreg
alcl2dr alcnow alcquit wgtadulbt yoyo10lb lcaldietb lfatdietb lsldietb fbdiet34b dbdiet34b lactdiet
othdiet walk walkmin walkspd hrdexb hrdexmin modexb modexmin mldexb mldexmin hrdex18 hrdex35 hrdex50
tepiwkb lepitotb msepikw xlmsepi phyactb sepiwk tminwk msminwk sminwk texpwkb walkexp smokingb hardexp
modexp mildexp avwkexp ffwkexp vfwkexp alcswwkb alcoholb;
KEEP id cigsdaysb smokyrbsb wgtadulbt lcaldietb lfatdietb lsldietb fbdiet34b dbdiet34b hrdexb
modexb mldexb hrdex18 hrdex35 hrdex50 smokingb tepiwkb lepitotb phyactb texpwkb alcswwkb alcoholb;
RUN;

/* thoughts and feelings */
DATA f37_ep_base_pub;
INFILE "&file_location.f37_ep_base_pub.dat" LRECL=474;
INPUT id f37days listen goodadvc takedr goodtime hlpchors share fun love livalnb livprt livchld
livsibl livrel livfrnds livoth pet dog cat bird fish othpet relgtime relstrn club hlpsick hlpsickt
nerves toomuch exclude coerce expctbst wrong hopefulb notmyway countgd moregood knwangry tellfeel
disappnt scenepub bother suppress apprvneg orders badluck truth lie honest unfair nocare trustno
frndsuse nohelp experts respect badsex lifequalb satlife genhelb hlthcly vigact modact liftgroc
stairs stair bending walklm walkblks walkblk bathing intsoc bodpainb painint lesswrkp lessaccp
lesskndp wrkdiffp lesswrke lessacce lesscare sickeasy hlthyany hlthwors hlthexcl intsoc2 fullpep
nervous dwndumps calm energy feltblue wournout happy tired eat dress inoutbed shower bloatingb
constipb nightswtb achesb brsttenb hotflashb diarrheab moodswngb nausea dizzy tiredb forgetb
hungry heartrac tremors heartbrn restless lowbackpb neckpainb skindryb headacheb clumsy trbsee vagitchb
concenb jntpainb nohunger hearloss swellhndb vagdryb upstomb urinpain cough vagdisb spousdie spousill
friendie monprob divorce frndiv chilcon majacc frnjob petdie feltdepb restslp enjlif cryspell
feltsad peopdis sad2wk sad2yrs sadmuch medsleep fallslp nap trbsleepb wakenght upearly backslp
snore qualslp hrsslp incontb frqincon noincon cghincon toincon slpincon othincon leakamt noprtct
minipad menspad diaper othprtct inconlmt incondis married actdly ambemot caregivl caregiv2
emolimit emowell enferat genhlth lfevent1 lfevent2 livalor nestemot optimism pain phylimit
physfunb pshtdep slpdstrb socfunc socstrn socsupp symptomb;
KEEP id livalnb hopefulb lifequalb genhelb bodpainb constipb diarrheab tiredb forgetb skindryb
upstomb feltdepb trbsleepb incontb physfunb symptomb bloatingb nightswtb achesb brsttenb
hotflashb moodswngb lowbackpb neckpainb headacheb vagitchb concenb jntpainb swellhndb vagdryb
vagdisb;
RUN;

/* hormone use */
DATA f43_ep_base_pub;
INFILE "&file_location.f43_ep_base_pub.dat" LRECL=99;
INPUT id f43days f43age recency oc ocagemin ocagemax octime tote totecat totemin totemax totestat totime
totp totpcat totpmin totpmax totpstat totptime toth tothcat tothmin tothmax tothstat tothtime pcycle;
KEEP id recency oc toth tothcat tothstat tothmin;
RUN;

/* medication use */
DATA f44_ct_pub;
INFILE "&file_location.f44_ct_pub.dat" LRECL=59;
INPUT id 1-6 day 8-12 vtyp 14 year 16-17 vclo 19 f44expc 21 medndc $ 23-35 adultdy 37-44 adulty 46-50
cort 52 tccode $ 54-59;
KEEP id day vtyp year vclo tccode medndc;
RUN;

/* medication use */
DATA f44ref_meds;
INFILE "&file_location.f44ref_meds.dat" LRECL=207;
INPUT medndc $ 1-13 medname $ 15-76 medabrv $ 78-104 medgener $ 106-160 medstren 170-178 meduom $ 180-192
medform $ 194-197 medrt $ 199-200 tccode $ 202-207 ;
KEEP medndc medgener medname;
RUN;

/* types of supplements */
DATA f45b_ep_base_pub;
INFILE "&file_location.f45b_ep_base_pub.dat" LRECL=72;
INPUT id f45days multib mvminb f45stres f45othcm f45combp tkvita viteb tkvitb1 tkvitb12 tkvitb2 tkvitb6
tkbeta tkbiot tkvitc tkcalc tkchrom tkcopp tkvitd tkfolic tkiron tkmagn tkmang tkmolyb tkniac tkpanto
tkphos tkpotas tkretin tksele tkzinc;
KEEP id multib mvminb viteb;
RUN;

DATA f60a_ep_base_pub;
INFILE "&file_location.f60a_ep_base_pub.dat" LRECL=338;
INPUT id f60days status f60grams f60enrgy f60enrgyj f60carb f60tsugr f60gtlc f60glac f60dietgi f60dietga
f60prot f60anmpr f60vegpr f60fat f60sfa f60mfa f60pfa f60tftot fruitbb vegebb f60fiber f60solfb
f60insfb f60chols f60water f60alc f60caff f60cbpct f60prpct f60ftpct f60sfpct f60mpfct f60pfpct;
KEEP id fruitbb vegebb;
RUN;

/* physical measurements */
DATA f80_ep_base_pub;
INFILE "&file_location.f80_ep_base_pub.dat" LRECL=90;
INPUT id f80days pulse30 systbp1 diasbp1 systbp2 diasbp2 heightx weightx waistx hipx syst systolb dias
diastolb bmib bmicxb whrx;
KEEP id systolb diastolb bmib bmicxb;

```

```

RUN;

/* breast exam */
DATA f84_ep_base_pub;
INFILE "&file_location.f84_ep_base_pub.dat" LRECL=57;
INPUT id f84days niprt niplt skinrt skinlt axilrt axillt brmassrt brmasslt brmoblrt brmobllt brsizert
      brsizelt massesrt masseslt brslfex brref cberefdy brfinrtb brfinltb cbeclcdy;
KEEP id brfinrtb brfinltb;
RUN;

/* mammogram */
DATA f85_ep_base_pub;
INFILE "&file_location.f85_ep_base_pub.dat" LRECL=42;
INPUT id f85days mammdy reviewdy rtbrst ltbrst furefer referdy mamrepet rfuresltb lfuresltb;
KEEP id rfuresltb lfuresltb;
RUN;

/*-----
  READ FOLLOWUP DATA
  -----*/

/* hrt management and safety interview, report of vaginal bleeding */
DATA f10_50_ep_fu_publ;
INFILE "&file_location_fu.f10_50_ep_fu_pub.dat" LRECL=115;
INPUT id day vtyp year hyst vagbleed heavybld vstartdy intermit bleednow vstopdy brsttend sevtend
      brstchnng corticos anticoag estrogen progest testost tamoxif endohyp hightrig trigl000 bldclot
      melanoma heartstk meningio brstcanc gallblad pancreas contmeds rtrneval gynnote physnote medchnng
      otheract pillfreq daysmiss allpills expsymp frgtpill frgtbotl tookbrk afraid famrcmnd mdrcmnd
      nopills othrmiss iap recontac;
RUN;

/* medical history update faints, falls, exams, tests, procedures */
DATA f33_ep_fu_pub;
INFILE "&file_location_fu.f33_ep_fu_pub.dat" LRECL=58;
INPUT id day vtyp year f33whom f33faint f33falls physexam eyeexam brstexam mammoqrm brstbpsy rctlexam
      hemoclt flexsig barium bpcheck cholchk ecg ptca outpdvt papsmr dandc endobpsy hysterec;
DROP f33whom hysterec;
RUN;

/* personal habits update */
DATA f35_ep_fu_pub;
INFILE "&file_location_fu.f35_ep_fu_pub.dat" LRECL=163;
INPUT id day vtyp year walk walkmin walkspd hrdex hrdexmin modex modexmin mldex mldexmin beerfreq beerserv
      winefreq wineserv liqrfreq liqrserv smoking cigsdays tepiwk lepitot msepikw xlmsepi phyact sepiwk
      tminwk msminwk sminwk texpwk walkexp avwkexp ffwkexp vfwkexp hardexp modexp mildexp alcswk;
KEEP id day vtyp year hrdex modex mldex smoking cigsdays tepiwk lepitot phyact texpwk alcswk;
RUN;

/* daily life */
DATA f38_ep_fu_pub;
INFILE "&file_location_fu.f38_ep_fu_pub.dat" LRECL=338;
INPUT id day vtyp year lifequal satlife genhel hlthcly vigact modact liftgroc stairs stair bending
      walklm walkblks walkblk bathing intsoc bodpain painint lesswrkp lessaccp lesskndp wrkdifff
      lesswrke lessacce lesscare sickeasy hlthyany hlthwors hlthexcl intsoc2 fullpep nervous dwndumps
      calm energy feltblue wornout happy tiredl eat dress inoutbed shower bloating constip nightswt aches
      brstten hotflash diarrhea moodswng nausea dizzy tired forget hungry heartrac tremors heartbrn restless
      lowbackp neckpain skindry headache clumsy trbsee vagitch concen jntpain nohunger hearloss swellhnd
      vagdry upstom urinpain cough vagdis spousdie spousill friendie monprob divorce frndiv chilcon
      majacc frnjob phyab verbab petdie feltdep restslep enjlif cryspell feltsad peopdis sad2wk sad2yrs
      sadmuch medsleep fallslp nap trbsleep wakenght uearly backslp snore qualslp hrsslp incont frqincon
      noincon cghincon toincon splincon othincon leakamt noprtct minipad menspad diaper
      othprtct inconlmt incondis married sexactiv satsex satfrqsx sexworry actdly emolimit emowell
      enerfat genhlth lfevent1 lfevent2 pain phylimit physfun pshtdep slpdstrb socfunc symptom;
KEEP id day vtyp year lifequal genhel bodpain constip diarrhea tired skindry upstom feltdep incont
      physfun symptom forget trbsleep bloating nightswt aches brstten hotflash moodswng lowbackp
      neckpain headache vagitch concen jntpain swellhnd vagdry vagdis;
RUN;

/* types of supplements */
DATA f45b_ep_fu_pub;
INFILE "&file_location_fu.f45b_ep_fu_pub.dat" LRECL=75;
INPUT id day vtyp year multi mvmin f45stres f45othcm f45combp tkvita vite tkvitb1 tkvitb2 tkvitb6
      tkbeta tkbiot tkvitc tkcalc tkchrom tkcopp tkvitd tkfolic tkiron tkmagn tkmang tkmolyb tkniac tkpanto
      tkphos tkpotas tkretin tkselen tkzinc;
KEEP id day vtyp year multi mvmin vite;
RUN;

/* physical measurements */
DATA f80_ep_fu_pub;
INFILE "&file_location_fu.f80_ep_fu_pub.dat" LRECL=97;
INPUT id day vtyp year pulse30 systbp1 diasbp1 systbp2 diasbp2 heightx weightx waistx hipx whexpect syst
      systol dias diastol bmi bmidx whrx;
KEEP id day vtyp year systol diastol bmi bmidx;
RUN;

/* clinical breast exam */
DATA f84_ep_fu_pub;
INFILE "&file_location_fu.f84_ep_fu_pub.dat" LRECL=66;
INPUT id day vtyp year cbeclcdy cbeBY cbeabnrm niprt niplt skinrt skinlt axilrt axillt brmassrt brmasslt
      brmoblrt brmobllt brsizert brsizelt massesrt masseslt brslfex brref cberefdy brfinrt brfinlt ;

```

```

KEEP id day vtyp year brfinrt brfinlt;
RUN;

/* mammogram */
DATA f85_ep_fu_pub;
INFILE "&file_location_fu.f85_ep_fu_pub.dat" LRECL=43;
INPUT id day vtyp year mammdy reviewdy rtbrst ltbrst furefer referdy mamrepet rfureslt lfureslt;
KEEP id day vtyp year rfureslt lfureslt;
RUN;

/* hrt medication adherence */
DATA ht_adh_ep_fu_pub;
INFILE "&file_location_fu.ht_adh_ep_fu_pub.dat" LRECL=41;
INPUT id vtyp year startdy enddy adhrate collect stophrt resumehrt lost dead openlabel ert2pert;
DROP vtyp;
RUN;

/* outcomes - cancer */
DATA ocan_ep_fu_pub;
INFILE "&file_location_fu.ocan_ep_fu_pub.dat" LRECL=341;
INPUT id brstc brstcdy insitu insitudy insdiag inshist insgrade inssize inssizecm invasive invady
invdiag invhist invhistic invgrade invnopair invright invleft invoneside invbilat invpaired
invsize invsizecm invinlv invnlymph invplymph invstage inverassay invprassay colotot colototdy
colon colondy rectal rectdy invcolo invcolody collocal collocdy coldetct coldetctdy colnodtct
colnodtctdy colregmet colregmetdy colsite colicd colhist colbehav colgrade colsize colinlv
colnlymph colplymph colstage endoca endocady epiendoca epiendocady enhist enbehavior engrade
ensize enlnodes enlymphve enlympnum enstage ovca ovcady ovcainv ovcainvdy epiinvov epiinvovdy
ovcalmp ovcalmpdy ovhist ovbehav ovgrade ovlateral ovszie ovlnodes ovlymphve ovlympnum ovstage
bladder bladderdy kidney kidneydy leuk leukdy lung lungdy lymphnh lymphnhdy mel meldy mult multdy
panc pancdy thy thydy lothca lothcady;
KEEP id brstc brstcdy insitu insitudy invasive invady colotot colototdy endoca endocady ovca
ovcady bladder bladderdy kidney kidneydy leuk leukdy lung lungdy lymphnh lymphnhdy mel
meldy mult multdy panc pancdy thy thydy lothca lothcady;
RUN;

/* outcomes - cvd */
DATA ocvd_ep_fu_pub;
INFILE "&file_location_fu.ocvd_ep_fu_pub.dat" LRECL=257;
INPUT id ecgpatt mi midy nofatmi nofatmidy nofatmi2 nofatmi2dy dsmi dsmidy psmi psmidy cordth cordthdy
chd chddy angina anginady anginadx1 anginadx2 anginadx3 anginadx4 anginadx5 anginadx6 anginadx7
confang confangdy acs acsdy acscat chf chfdy revasc revascdy cad caddy caddiag cadhsp cadangio
cadsurg herscad pad paddy paddiag loextrem aaa hersloex pvd pvddy dvt dvt2 dvt1 dvtldy dvt2
dvt2dy dvt2schg dvtveno dvtimpd dvtinolv dvtisotop dvtreprt peworkup pe pedy pel peldy pe2 pe2dy
pedschg pelngscan peangio pesuggest peother vte vtedy vte1 vteldy vte2 vte2dy stroke strokedy
diagnosis oxford toast glasgow totchd totchddy;
KEEP id mi midy nofatmi nofatmidy nofatmi2 nofatmi2dy dsmi dsmidy cordth cordthdy chd chddy
angina anginady confang confangdy acs acsdy chf chfdy revasc revascdy cad caddy pad paddy
pvd pvddy dvt dvt2dy pe pedy vte vtedy stroke strokedy totchd totchddy;
RUN;

/* outcomes - fractures */
DATA ofrc_ep_fu_pub;
INFILE "&file_location_fu.ofrc_ep_fu_pub.dat" LRECL=194;
INPUT id hip hipdy hipfxtype hipfxside hippathfx femneck femneckdy intertroc intertrocdy elbow elbowdy erad
eraddy elbownos elbownosdy foot footdy hand handdy pat patdy knee kneedy larmwadj larmwdy rad raddy
carpal carpaldy lowleg lowlegdy ankle ankledy pelvis pelvisdy spineadj spinedy sacc saccdy hummue
hummuedy humms hummsdy clavicle clavicle dy scap scapdy upleg uplegdy unkfx unkfxdy ofracadj othfrcdy
fracadj fracdy;
KEEP id fracadj fracdy;
RUN;

/* outcomes - other */
DATA ooth_ep_fu_pub;
INFILE "&file_location_fu.ooth_ep_fu_pub.dat" LRECL=205;
INPUT id cataract catarady glaucoma glaucdy osteop osteopdy osteoart ostartdy rheumart rheumdy lupus
lupusdy diabpill diabpdy diabshot diabsdy diabslf diabslfdy hyppill hyppildy heart heartdy
cathhosp cthslfy othheart othhrtdy ampirc ampircdy intpoly intplydy kidstone kidstndy
choleitis choledy stone stonedy galldis galldy cholectmy cholepdy othproc othpdy gallproc
gallpdy gallbladder gallblady hysterectomy hystdy ooph oophdy hospever hospdy hospnum death
deathdy deathcat lastcont;
DROP diabpill diabpdy diabshot diabsdy;
RUN;

/*****
MERGE BASELINE & OUTCOME DATA
*****/

%LET datasets = dem_ep_base_pub f2_ep_base_pub f20_ep_base_pub f30_ep_base_pub f31_ep_base_pub
f32_ep_base_pub f34_ep_base_pub f37_ep_base_pub f43_ep_base_pub f45b_ep_base_pub f60a_ep_base_pub
f80_ep_base_pub f84_ep_base_pub f85_ep_base_pub ocan_ep_fu_pub ocvd_ep_fu_pub ofrc_ep_fu_pub ooth_ep_fu_pub;

%DO I = 1 %TO 18;
%LET WORD = %SCAN(&DATASETS,%EVAL(&I),%STR( ));
PROC SORT DATA = &WORD ;
BY id ;
RUN;

%END ;

```

```

DATA whi.baseline_bc;
MERGE &DATASETS;
BY id;
RUN;

PROC DATASETS LIBRARY = work NOLIST ;
DELETE &DATASETS ;
QUIT;

/*-----
   CREATE CONSISTENT DEFINITION OF YEAR BASED ON VALUE OF DAY
-----*/
%LET datasets = f44_ct_pub f10_50_ep_fu_pub1 f33_ep_fu_pub f35_ep_fu_pub f38_ep_fu_pub f45b_ep_fu_pub
               f80_ep_fu_pub f84_ep_fu_pub f85_ep_fu_pub;
%DO i = 1 %TO 9;
  %LET datain = %SCAN(&DATASETS,%EVAL(&i),%STR( ));
  DATA &datain ;
  SET &datain (DROP=year);
  IF day <= 0 THEN year = 1;
  ELSE year = CEIL(day/365.25);
RUN;
%END;

/*****
   CREATE MEDICATION USE DATA (BASELINE & FOLLOWUP)
*****/

/*-----
   EXTRACT DATA FOR WHI E+P TRIAL PARTICIPANTS
-----*/
PROC SORT DATA=f44ref_meds;
BY mednrc;
RUN;

PROC SORT DATA=f44_ct_pub;
BY mednrc;
RUN;

DATA f44_ct_pub1;
MERGE f44_ct_pub f44ref_meds;
BY mednrc;
RUN;

DATA id;
SET whi.baseline_bc(KEEP=id lastcont chddy deathdy);
RUN;

DATA f44_ep_base_pub;
SET f44_ct_pub1( WHERE = ( vtyp = 1)); /* vtyp=1: screening visit */
RUN;

DATA f44_ep_fu_pub;
SET f44_ct_pub1( WHERE = ( vtyp NE 1 AND vclo=1)); /* closest to visit within visit type and year */
RUN;

DATA f44_ep_base_pub;
SET f44_ep_base_pub(KEEP=id tccode medgener medname);
RUN;

DATA f44_ep_fu_pub;
SET f44_ep_fu_pub (KEEP=id year day vtyp tccode medgener medname);
RUN;

/* keep only those observations for subjects in baseline_bc */

PROC SQL;
CREATE TABLE med_ep_base AS
SELECT M.*, S.*
FROM id M LEFT JOIN f44_ep_base_pub S
ON M.id=S.id;
QUIT;

PROC SQL;
CREATE TABLE med_ep_ful AS
SELECT M.*, S.*
FROM id M LEFT JOIN f44_ep_fu_pub S
ON M.id=S.id;
QUIT;

DATA med_ep_fu;
SET med_ep_ful;
IF lastcont=. THEN DO;
  IF chddy NE . AND deathdy NE . THEN lastcont=chddy;
  ELSE IF chddy NE . AND deathdy = . THEN lastcont=chddy;

```

```

ELSE IF chddy = . AND deathdy NE . THEN lastcont=deathdy;
ELSE lastcont=5.6*365.25;
END;
IF year=. OR year>CEIL(lastcont/365.25) THEN DELETE;
DROP lastcont chddy deathdy;
RUN;

```

```

/*-----
CREATE INDICATORS FOR MEDICATION USE (BASELINE)
-----*/

```

```

PROC TRANSPOSE DATA=med_ep_base OUT=med_base1 PREFIX=tccode;
BY id;
VAR tccode;
RUN;

```

```

DATA med_base1;
SET med_base1;
statinb=0;
sermb=0;
bisphob=0;
hilipidb=0;
hypertensionb=0;
diabetesb=0;
corticosteroidb=0;
anticoagulantb=0;
ARRAY a(*) tccode1--tccode24;
DO i=1 TO DIM(a);
IF a(i) IN ('394000','394099') THEN statinb=1;
IF a(i) IN ('300530') THEN sermb=1;
IF a(i) IN ('300420') THEN bisphob=1;
IF a(i) >= 390000 AND a(i) < 400000 THEN hilipidb=1;
IF a(i) >= 360000 AND a(i) < 380000 THEN hypertensionb=1;
IF a(i) >= 270000 AND a(i) < 280000 THEN diabetesb=1;
IF a(i) >= 220000 AND a(i) < 230000 THEN corticosteroidb=1;
IF a(i) >= 830000 AND a(i) < 840000 THEN anticoagulantb=1;
END;
DROP tccode1-tccode24 _name_ i;
RUN;

```

```

PROC TRANSPOSE DATA=med_ep_base OUT=med_base2 PREFIX=medgener;
BY id;
VAR medgener;
RUN;

```

```

DATA med_base2;
SET med_base2;
aspirinb=0;
tamoxifenb=0;
testosteroneb=0;
ARRAY a(*) medgener1--medgener24;
DO i=1 TO DIM(a);
IF INDEX(a(i),"ASPIRIN") THEN aspirinb=1;
IF INDEX(a(i),"TAMOXIFEN") THEN tamoxifenb=1;
IF INDEX(a(i),"TESTOSTERONE") THEN testosteroneb=1;
END;
DROP medgener1-medgener24 _name_ i;
RUN;

```

```

PROC TRANSPOSE DATA=med_ep_base OUT=med_base3 PREFIX=medname;
BY id;
VAR medname;
RUN;

```

```

DATA med_base3;
SET med_base3;
estprob=0;
ARRAY a(*) medname1--medname24;
DO i=1 TO DIM(a);
IF INDEX(a(i),"PREMPRO") then estprob=1;
END;
DROP medname1-medname24 _name_ i;
RUN;

```

```

DATA whi.baseline_bc;
MERGE whi.baseline_bc med_base1 med_base2 med_base3 ;
BY id;
RUN;

```

```

/*-----
CREATE INDICATORS FOR MEDICATION USE (FOLLOWUP)
-----*/

```

```

PROC SORT DATA=med_ep_fu;
BY id year vtyp day;
RUN;

```

```

PROC TRANSPOSE DATA=med_ep_fu OUT=med_ful PREFIX=tccode;
BY id year;
var tccode;

```

```

RUN;

DATA med_ful;
SET med_ful;
statin=0;
serm=0;
bispho=0;
hilipid=0;
hypertension=0;
diabetes=0;
corticosteroid=0;
anticoagulant=0;
ARRAY a(*) tccodel--tccode31;
DO i=1 TO DIM(a);
  IF a(i) IN ('394000','394099') THEN statin=1;
  IF a(i) IN ('300530') THEN serm=1;
  IF a(i) IN ('300420') THEN bispho=1;
  IF a(i) >= 390000 AND a(i) < 400000 THEN hilipid=1;
  IF a(i) >= 360000 AND a(i) < 380000 THEN hypertension=1;
  IF a(i) >= 270000 AND a(i) < 280000 THEN diabetes=1;
  IF a(i) >= 220000 AND a(i) < 230000 THEN corticosteroid=1;
  IF a(i) >= 830000 AND a(i) < 840000 THEN anticoagulant=1;
END;
DROP tccodel-tccode31 _name_ i;
RUN;

PROC TRANSPOSE DATA=med_ep_fu OUT=med_fu2 PREFIX=medgener;
BY id year;
VAR medgener;
RUN;

DATA med_fu2;
SET med_fu2;
aspirin=0;
tamoxifen=0;
testosterone=0;
ARRAY a(*) medgener1--medgener31;
DO i=1 TO DIM(a);
  IF INDEX(a(i),"ASPIRIN") then aspirin=1;
  IF INDEX(a(i),"TAMOXIFEN") then tamoxifen=1;
  IF INDEX(a(i),"TESTOSTERONE") then testosterone=1;
END;
DROP medgener1-medgener31 _name_ i;
RUN;

PROC TRANSPOSE DATA=med_ep_fu OUT=med_fu3 PREFIX=medname;
BY id year;
VAR medname;
RUN;

DATA med_fu3;
SET med_fu3;
estpro=0;
ARRAY a(*) medname1--medname31;
DO i=1 TO DIM(a);
  IF SUBSTR(a(i),1,8)="PREMPRO" THEN estpro=1;
END;
DROP medname1-medname31 _name_ i;
RUN;

DATA med_fu;
MERGE med_ful med_fu2 med_fu3;
BY id year;
RUN;

PROC DATASETS LIBRARY = work NOLIST ;
DELETE med_ful med_fu2 med_fu3 med_base1 med_base2 med_base3 ;
QUIT;

```

```

/*-----
  RECODE BASELINE VARIABLES
-----*/
DATA whi.baseline_bc;
SET whi.baseline_bc;
/* dem_ep_base_pub */
IF ethnic IN (.,1,2,8) THEN ethnic=8;
IF educ1 = . THEN educ=9;
IF . < educ1 <= 5 THEN educ=0;
ELSE IF 5 < educ1 <= 7 THEN educ=1;
ELSE IF educ1 > 7 THEN educ=2;
IF . < incomel <= 2 THEN income=1;
ELSE IF 2 < incomel <= 4 THEN income=2;
ELSE IF 4 < incomel <= 8 THEN income=3;
ELSE IF incomel IN (.,9) THEN income=9;

/* f2_ep_base_pub */
IF diabtrtb=. THEN diabtrtb=0;

/* f20_ep_base_pub */
IF marital=. THEN marital=9;

```



```

IF marital=5 THEN marital=4;
IF careprov=. THEN careprov=0;
IF mammob=. THEN mammob=0;
IF nomam2yr=. THEN nomam2yr=0;
IF papsmearb IN (.,9) THEN papsmearb=0;
IF abnpap3y=. THEN abnpap3y=0;
IF cervdys=. THEN cervdys=0;
IF endoasP=. THEN endoasP=0;
IF timelsts=. THEN timelsts=0;
IF anyins=. THEN anyins=0;

```

```

/* f30_ep_base_pub */
ARRAY a(*) hosp2y--revascb;
DO b=1 TO DIM(a);
  IF a(b)=. THEN a(b)=0;
END;

```

```

IF numfallsb=0 THEN numfallsb=0;
IF numfallsb>0 THEN numfallsb=1;

```

```

/* f31_ep_base_pub */
IF menarcheb = . THEN menarche=9;
IF . < menarcheb <= 3 THEN menarche=0;
IF menarcheb in (4,5) THEN menarche=1;
IF menarcheb >= 6 THEN menarche=2;
IF menpsympb=. THEN menpsympb=0;
IF nedlaspb=. THEN nedlaspb=0;
IF brstbiopb=. THEN brstbiopb=0;
IF gravid=. THEN gravid=9;
parity=parity+1;
IF parity=. THEN parity=9;

```

```

IF brstfeed NE 1 THEN brstfed=0;
IF brstfeed = 1 THEN DO;
  IF brstfdn = . THEN brstfed=1;
  IF brstfdn IN (1,2) THEN brstfed=1;
  IF brstfdn >= 3 THEN brstfed=2;
END;
IF agefbir=. THEN agefbir=9;
IF boophb=. THEN boophb=0;
IF brstdisb=. THEN brstdisb=9;

```

```

/* f32_ep_base_pub */
ARRAY c(*) diabrel--colorel;
DO d=1 TO DIM(c);
  IF c(d) IN (.,9) THEN c(d)=0;
END;

```

```

IF brcamom in (1,2,3) THEN brcamom=1;
ELSE brcamom=0;
IF brcasis1 in (1,2,3) THEN brcasis1=1;
ELSE brcasis1=0;
IF brcasis2 in (1,2,3) THEN brcasis2=1;
ELSE brcasis2=0;
IF brcasis3 in (1,2,3) THEN brcasis3=1;
ELSE brcasis3=0;
IF brcadaul in (1,2,3) THEN brcadaul=1;
ELSE brcadaul=0;
IF brcadau2 in (1,2,3) THEN brcadau2=1;
ELSE brcadau2=0;
IF brcadau3 in (1,2,3) THEN brcadau3=1;
ELSE brcadau3=0;
bctnum=SUM(brcamom,brcasis1,brcasis2,brcasis3,brcadaul,brcadau2,brcadau3);
IF bctnum>=2 THEN bctnum=2;

```

```

/* f34_ep_base_pub */
IF texpwkb = . THEN metb=9;
IF texpwkb=0 THEN metb=0;
IF 0.00<texpwkb<= 4.25 THEN metb=1;
IF 4.25<texpwkb<=10.00 THEN metb=2;
IF 10.00<texpwkb<=19.50 THEN metb=3;
IF texpwkb> 19.50 THEN metb=4;
IF alcoholb=. THEN alcoholb=9;
IF smokingb=. THEN smokingb=9;

```

```

/* f37_ep_base_pub */
IF hopefulb=. THEN hopefulb=9;
IF genhelb=. THEN genhelb=9;
IF bodpainb=. THEN bodpainb=9;
IF physfunb=. THEN physfunb=0;
IF physfunb<=75 THEN physfunb=1;
IF 75<physfunb<=90 THEN physfunb=2;
IF 90<physfunb<=95 THEN physfunb=3;
IF physfunb> 95 THEN physfunb=4;

```

```

IF constipb IN (.,0) THEN constipb=0;
ELSE constipb=1;
IF diarrheab IN (.,0) THEN diarrheab=0;
ELSE diarrheab=1;
IF upstomb IN (.,0) THEN upstomb=0;
ELSE upstomb=1;
IF feltdepb IN (.,0) THEN feltdepb=0;

```

```

ELSE feltdepb=1;
IF incontb = . THEN incontb=0;
IF skindryb IN (.,0) THEN skindryb=0;
IF tiredb IN (.,0) THEN tiredb=0;
IF forgetb IN (.,0) THEN forgetb=0;
IF trbsleepb IN (.,0) THEN trbsleepb=0;
IF symptomb = . THEN symptomb=0;

IF bloatingb IN (.,0) THEN bloatingb=0;
IF nightswtb IN (.,0) THEN nightswtb=0;
IF achesb IN (.,0) THEN achesb=0;
IF brsttenb IN (.,0) THEN DO;
    brsttenb=0;
    brsttendb=0;
    sevtendb=0;
END;
IF brsttenb=1 THEN DO;
    brsttendb=1;
    sevtendb=1;
END;
IF brsttenb IN (2,3) THEN DO;
    brsttendb=1;
    sevtendb=2;
END;
IF hotflashb IN (.,0) THEN hotflashb=0;
IF moodswngb IN (.,0) THEN moodswngb=0;
IF lowbackpb IN (.,0) THEN lowbackpb=0;
IF neckpainb IN (.,0) THEN neckpainb=0;
IF headacheb IN (.,0) THEN headacheb=0;
IF vagitchb IN (.,0) THEN vagitchb=0;
IF concenb IN (.,0) THEN concenb=0;
IF jntpainb IN (.,0) THEN jntpainb=0;
IF swellhndb IN (.,0) THEN swellhndb=0;
IF vagdryb IN (.,0) THEN vagdryb=0;
IF vagdisb IN (.,0) THEN vagdisb=0;

/* f43_ep_base_pub */
IF recency=. THEN recency=0;
IF tothcat=. THEN tothcat=0;
IF tothstat=. THEN tothstat=0;

/* f45b_ep_base_pub */
IF multib=1 OR mvminb=1 THEN multib=1;
ELSE multib=0;
IF viteb=1 THEN viteb=1;
ELSE viteb=0;

/* f60a_ep_base_pub */
IF fruitb=. THEN fruitb=0;
IF vegeb=. THEN vegeb=0;

/* f80_ep_base_pub */
IF bmicxb=. THEN bmicxb=9;

/* med_base */
IF statinb=. THEN statinb=0;
IF sermb=. THEN sermb=0;
IF bisphob=. THEN bisphob=0;
IF hilipidb=. THEN hilipidb=0;
IF hypertensionb=. THEN hypertensionb=0;
IF diabetesb=. THEN diabetesb=0;
IF corticosteroidb=. THEN corticosteroidb=0;
IF anticoagulantb=. THEN anticoagulantb=0;
IF aspirinb=. THEN aspirinb=0;
IF tamoxifenb=. THEN tamoxifenb=0;
IF testosteroneb=. THEN testosteroneb=0;
IF estprob=. THEN estprob=0;
IF sermb=0 AND tamoxifenb=1 THEN sermb=1;

/* f84_ep_base_pub */
IF brfinrtb= . OR brfinltb= . THEN brfinb=9;
IF brfinrtb= 0 & brfinltb= 0 THEN brfinb=0;
IF brfinrtb= 1 & brfinltb<=1 THEN brfinb=1;
IF brfinrtb<=1 & brfinltb= 1 THEN brfinb=1;
IF brfinrtb> 1 OR brfinltb> 1 THEN brfinb=2;

/* f85_ep_base_pub */
IF rfuresltb= . OR lfuresltb= . THEN mafinb=9;
IF rfuresltb= 0 & lfuresltb= 0 THEN mafinb=0;
IF rfuresltb= 1 & lfuresltb<=1 THEN mafinb=1;
IF rfuresltb<=1 & lfuresltb= 1 THEN mafinb=1;
IF rfuresltb> 1 OR lfuresltb> 1 THEN mafinb=2;

IF brfinb IN (0,9) & mafinb IN (0,9) THEN brmafinb=0;
IF brfinb = 1 & mafinb IN (0,1,9) THEN brmafinb=1;
IF brfinb IN (0,1,9) & mafinb = 1 THEN brmafinb=1;
IF brfinb = 2 OR mafinb = 2 THEN brmafinb=2;
IF brfinb = 9 & mafinb = 9 THEN brmafinb=9;

IF mib=1 OR strokeb=1 OR anginab=1 OR revascb=1 OR carotidb=1 OR tiab=1 OR padb=1 THEN chdb=1;
ELSE chdb=0;

```

```

IF htnttrtb IN (1,2) OR (systolb=3 AND diastolb=2) OR hypertensionb=1 THEN htnb=1;
ELSE htnb=0;
IF midad=1 OR mimom IN (1,2) THEN mIFam=1;
ELSE mIFam=0;
IF hicholb=1 OR hilipidb=1 THEN hicholb=1;
ELSE hicholb=0;
IF diabtrtb=1 OR diabetesb=1 THEN diabtrtb=1;
ELSE diabtrtb=0;

cardiob = (SUM(SUM(dvtb,peb,arrestb,chfb,cathb,afb,aneurysmb),cvdb)>0);
cancerb = (SUM(canb,brcab,colonb,cervcab)>0);

menopauseb=SUM(bloatingb,nightswtb,achesb,brsttenb,hotflashb,moodswngb,lowbackpb,neckpainb,headacheb,
vagitcb,concenb,jntpainb,swellhndb,vagdryb,vagdisb);
IF 0<=menopauseb<=4 THEN menosymb=0;
IF 4< menopauseb<=6 THEN menosymb=1;
IF 6< menopauseb<=10 THEN menosymb=2;
IF menopauseb> 10 THEN menosymb=3;

IF hotflashb=0 AND nightswtb=0 THEN vasomotorb=0;
ELSE IF hotflashb=0 AND nightswtb=1 THEN vasomotorb=1;
ELSE IF hotflashb=1 AND nightswtb=0 THEN vasomotorb=1;
ELSE IF hotflashb=1 AND nightswtb=1 THEN vasomotorb=1;
ELSE vasomotorb=2;

IF 0 <=fruitbb<0.88219 THEN fruitb=0;
IF 0.88219<=fruitbb<1.50411 THEN fruitb=1;
IF 1.50411<=fruitbb<2.49315 THEN fruitb=2;
IF 2.49315<=fruitbb THEN fruitb=3;

IF 0 <=vegebb<1.18630 THEN vegeb=0;
IF 1.18630<=vegebb<1.84110 THEN vegeb=1;
IF 1.84110<=vegebb<2.77808 THEN vegeb=2;
IF 2.77808<=vegebb THEN vegeb=3;

preventb1=SUM(mammob,papsmearb,endoasp,colnscpyb,pcolonrmb,hemocculb,nedlaspb,boophb);
IF preventb1 IN (0,1,2) THEN preventb=0;
ELSE IF preventb1 IN (3,4) THEN preventb=1;
ELSE IF preventb1 >= 5 THEN preventb=2;

/*-----
DEFINE YEAR DURING WHICH INVASIVE BREAST CANCER OR DEATH OCCURRED
-----*/
IF invasive=1 THEN invayear=CEIL(invady/365.25);
ELSE invayear=.;
IF death=1 THEN dyear=CEIL(deathdy/365.25);
ELSE dyear=.;

/*-----
DEFINE YEARS SINCE MENOPAUSE (MENOAGE)
-----*/
IF oophb IN (0,1,3,4,9) THEN biooph=0;
IF oophb=2 THEN biooph=1;
IF oophb=. THEN biooph=.;

IF biooph=1 THEN bioopha=oophab;
IF biooph IN (.,0) THEN bioopha=.;

IF bioopha=1 THEN bioocat=30;
IF bioopha=2 THEN bioocat=32;
IF bioopha=3 THEN bioocat=37;
IF bioopha=4 THEN bioocat=42;
IF bioopha=5 THEN bioocat=47;
IF bioopha=6 THEN bioocat=52;
IF bioopha=7 THEN bioocat=57;
IF bioopha=8 THEN bioocat=60;
IF bioopha=. THEN bioocat=.;

IF 1<=hystageb<=3 THEN agehyst=1;
IF 4<=hystageb<=5 THEN agehyst=2;
IF hystageb= 6 THEN agehyst=3;
IF hystageb> 6 THEN agehyst=4;
IF hystageb= . THEN agehyst=.;
IF hystageb= 8 AND 50<=age<=59 THEN agehyst=.;

IF agehyst=1 THEN hystcat=40;
IF agehyst=2 THEN hystcat=45;
IF agehyst=3 THEN hystcat=52;
IF agehyst=4 THEN hystcat=55;
IF agehyst=. THEN hystcat=.;

meno=MIN(anymensa,bioocat,tothmin);

IF hystb=1 & oophb IN (0,1,3,4,9) THEN meno=MIN(tothmin,menpsyaf);
IF hystb=1 & hystageb >= 3 & oophb IN (0,1,3,4,9) & menpsyaf=. & (tothmin=. OR tothmin>60) THEN meno=hystcat;
IF meno>60 THEN meno=60;
IF hystb=0 & anymensa=. & bioocat=. & tothmin=. THEN meno=.;
IF hystb=1 & oophb IN (0,1,3,4,9) & agehyst<3 & tothmin=. & menpsyaf=. THEN meno=.;

```

```

IF .< age-meno<10 THEN menoage=0;
ELSE IF 10<=age-meno<20 THEN menoage=1;
ELSE IF 20<=age-meno THEN menoage=2;
ELSE menoage=9;

DROP htntrtb hypertensio b systolb diastolb diabetesb midad mimom cervcab abnpap3y cervdys cancfrel
cancmrel cervrel endorel ovarrel menopauseb bloatingb achesb brsttenb moodswngb lowbackpb neckpainb
headacheb vagitchb concenb jntpainb swellhndb vagdryb vagdisb preventbl oophb biooph oophab bioopha
bioocat hystageb agehyst hystcat anymensa tothmin menpsyaf b d incomel brstfeed brcamom
brcasis1-brcasis3 brcadaul-brcadau3 mvminb tamoxifenb brfinrtb brfinltb rfuresltb lfuresltb;

RUN;

/*-----
INFO THAT WILL BE USED TO CALCULATE DOSE
-----*/

DATA adh_a;
SET f10_50_ep_fu_pub1 (KEEP=id year day vtyp pillfreq WHERE = (vtyp in (2,3)));
RUN;

PROC SORT DATA=adh_a;
BY id year vtyp day;
RUN;

DATA adh_a;
SET adh_a;
BY id year vtyp;
IF LAST.vtyp = 0 then delete ;
RUN;

PROC TRANSPOSE DATA=adh_a OUT=adh_c PREFIX=pillfreq;
BY id year;
VAR pillfreq;
RUN;

PROC SQL; CREATE TABLE f10_50_ep_fu_pub AS
SELECT O.*, L.pillfreq1, L.pillfreq2
FROM f10_50_ep_fu_pub1 O LEFT JOIN adh_c L
ON O.id=L.id AND o.year=l.year;
QUIT;

/*-----
REPLACE MISSING VALUE IN A GIVEN VISIT WITH THE MOST RECENT VISIT W/O MISSING VALUE
-----*/

/* f10_50_ep_fu_pub */
PROC SORT DATA=f10_50_ep_fu_pub;
BY id year day &order vtyp ; /* try to use multiple day values using the smallest vtyp value */
RUN;

DATA f10_50_ep_fu_pub;
SET f10_50_ep_fu_pub;
BY id year day;
IF first.day;
RUN;

DATA a;
SET f10_50_ep_fu_pub;
BY id;
ARRAY a(*) hyst vagbleed heavybld vstartdy intermit bleednow vstopdy brsttend sevtend brstchn g corticos
anticoag testost tamoxif endohyp hightrig trig1000 bldclot melanoma heartstk meningio brstcanc
gallblad pancreas contmeds rtrneval gynnote physnote medchn g otheract pillfreq daysmiss allpills
expsymp frgtpill frgtbotl tookbrk afraid famrcmnd mdrcmnd nopills othrmiss iap recontac;

ARRAY b(*) ahyst avagbleed aheavybld avstartdy aintermit ableednow avstopdy abrsttend asevtend abrstchn g
acorticos aanticoag atestost atamoxif aendohyp ahightrig atrig1000 abldclot amelanoma aheartstk
ameningio abrstcanc agallblad apancreas acontmeds artrneval agynnote aphysnote amedchn g aotheract
apillfreq adaysmiss aallpills aexpsymp afrgtpill afrgtbotl atookbrk aafraid afamrcmnd amdrcmnd
anopills aothrmiss aiap arecontac;

RETAIN ahyst avagbleed aheavybld avstartdy aintermit ableednow avstopdy abrsttend asevtend abrstchn g
acorticos aanticoag atestost atamoxif aendohyp ahightrig atrig1000 abldclot amelanoma aheartstk
ameningio abrstcanc agallblad apancreas acontmeds artrneval agynnote aphysnote amedchn g aotheract
apillfreq adaysmiss aallpills aexpsymp afrgtpill afrgtbotl atookbrk aafraid afamrcmnd amdrcmnd
anopills aothrmiss aiap arecontac ;

IF FIRST.id THEN DO;
DO i=1 TO DIM(a);
b(i)=a(i);
END;
END;
ELSE DO;
DO i=1 TO DIM(a);
IF a(i) NE . THEN b(i)=a(i);
END;
END;

DROP hyst vagbleed heavybld vstartdy intermit bleednow vstopdy brsttend sevtend brstchn g corticos anticoag
testost tamoxif endohyp hightrig trig1000 bldclot melanoma heartstk meningio brstcanc gallblad pancreas
contmeds rtrneval gynnote physnote medchn g otheract pillfreq daysmiss allpills expsymp frgtpill frgtbotl

```

```

tookbrk afraid famrcmnd mdrcmnd nopills othrmiss iap recontac i;
RUN;

PROC DATASETS LIBRARY = work NOLIST ;
MODIFY a;
RENAME ahyst=hyst avagbleed=vagbleed aheavybld=heavybld avstartdy=vstartdy aintermit=intermit ableednow=bleednow
avstopdy=vstopdy abrsttend=brsttend asevtend=sevtend abrstchnge=brstchnge acorticos=corticos aanticoag=anticoag
atestost=testost atamoxif=tamoxif aendohyp=endohyp ahightrig=hightrig atrigl000=trigl000 abldclot=bldclot
amelanoma=melanoma aheartstk=heartstk ameningio=meningio abrstcanc=brstcanc agallblad=gallblad apancreas=pancreas
acontmeds=contmeds artrneval=rtrneval agynnote=gynnote aphysnote=physnote amedchnge=medchnge aotheract=otheract
apillfreq=pillfreq adaysmiss=daysmiss aallpills=allpills aexpsymp=expsymp afrgtpill=frgtpill afrgtbotl=frgtbotl
atookbrk=tookbrk aafraid=afraid afamrcmnd=famrcmnd amdrcmnd=mdrcmnd anopills=nopills aothrmiss=othrmiss
aiap=iap arecontac=recontac;

QUIT;

/* f33_ep_fu_pub */
PROC SORT DATA=f33_ep_fu_pub OUT=b;
BY id day &order vtyp;
RUN;

DATA b;
SET b;
BY id day ;
IF first.day;
RUN;

DATA b;
SET b;
BY id;
ARRAY c(*) f33faint f33falls physexam eyeexam brstexam mammogrm brstbpsy rctlexam hemoclt flexsig barium bpcheck
cholchk ecg ptca outpdvt papsmr dandc endobpsy;
ARRAY d(*) af33faint af33falls aphysexam aeyeexam abrstexam amammogrm abrstbpsy arctlexam ahemoclt aflexsig abarium
abpcheck acholchk aecg aptca aoutpdvt apapsmr adandc aendobpsy;
RETAIN af33faint af33falls aphysexam aeyeexam abrstexam amammogrm abrstbpsy arctlexam ahemoclt aflexsig abarium
abpcheck acholchk aecg aptca aoutpdvt apapsmr adandc aendobpsy;
IF FIRST.id THEN DO;
DO i=1 TO DIM(c);
d(i)=c(i);
END;
END;
ELSE DO;
DO i=1 TO DIM(c);
IF c(i) NE . THEN d(i)=c(i);
END;
END;
END;
DROP f33faint f33falls physexam eyeexam brstexam mammogrm brstbpsy rctlexam hemoclt flexsig barium bpcheck cholchk
ecg ptca outpdvt papsmr dandc endobpsy i;
RUN;

PROC DATASETS LIBRARY = work NOLIST;
MODIFY b;
RENAME af33faint=f33faint af33falls=f33falls aphysexam=physexam aeyeexam=eyeexam abrstexam=brstexam amammogrm=mammogrm
abrstbpsy=brstbpsy arctlexam=rctlexam ahemoclt=hemoclt aflexsig=flexsig abarium=barium abpcheck=bpcheck
acholchk=cholchk aecg=ecg aptca=ptca aoutpdvt=outpdvt apapsmr=papsmr adandc=dandc aendobpsy=endobpsy;
QUIT;

/*-----
EXTRACT ANNUAL VISIT
-----*/
PROC SORT DATA=a;
BY id year day &order vtyp;
RUN;

DATA a ;
SET a ;
BY id year day ;
IF FIRST.day ;
RUN;

DATA f10_50;
SET a;
BY id year day;
IF LAST.year=1;
RUN;

PROC SORT DATA=b;
BY id year day &order vtyp;
RUN;

DATA b;
SET b;
BY id year day;
IF FIRST.day;
RUN;

DATA f33;
SET b;
BY id year day;
IF LAST.year=1;

```

RUN;

```
PROC SORT DATA=ht_adh_ep_fu_pub;
BY id year;
RUN;
```

```
PROC SORT DATA=f10_50;
BY id year day &order vtyp;
RUN;
```

```
DATA f10_50;
SET f10_50;
BY id year day;
IF FIRST.day;
RUN;
```

```
PROC SORT DATA=f33;
BY id year day &order vtyp;
RUN;
```

```
DATA f33;
SET f33;
BY id year day ;
IF FIRST.day ;
RUN;
```

```
/* f35_ep_fu_pub */
PROC SORT DATA=f35_ep_fu_pub;
BY id year day &order vtyp;
RUN;
```

```
DATA f35_ep_fu_pub;
SET f35_ep_fu_pub;
BY id year day;
IF FIRST.day;
RUN;
```

```
%LET varlist = hrdex modex mldex smoking cigsdays tepiwk lepitot phyact texpwk alcswk;
%LET n = 10 ;
%LET avarlist = ;
%DO i = 1 %TO &n ;
    %LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
    %LET avarlist = &avarlist a&var_i ;
%END;
```

```
DATA f35_ep_fu_pub (DROP = &avarlist );
SET f35_ep_fu_pub (RENAME = ( %DO i = 1 %TO &n ;
    %LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
    %LET avarlist = %SCAN(&avarlist, %EVAL(&i),%STR( ));
    &var_i = &avarlist
    %END;
));
```

```
BY id year ;
ARRAY c(*) &avarlist ;
ARRAY d(*) &varlist ;
RETAIN &varlist ;
```

```
IF FIRST.id THEN DO;
    DO i=1 TO DIM(c);
        d(i)=c(i);
    END;
END;
ELSE DO;
    DO i=1 TO DIM(c);
        IF c(i) NE . THEN d(i)=c(i);
    END;
END;
```

```
END;
IF LAST.year THEN OUTPUT;
RUN;
```

```
/* f38_ep_fu_pub */
PROC SORT DATA=f38_ep_fu_pub;
BY id year day &order vtyp;
RUN;
```

```
DATA f38_ep_fu_pub ;
SET f38_ep_fu_pub;
BY id year day;
IF FIRST.day ;
```

```
%LET varlist = lifequal genhel bodpain constip diarrhea tired skindry upstom feltdep incont
    physfun symptom forget trbsleep bloating nightswt aches brstten hotflash moodswng
    lowback neckpain headache vagitch concen jntpain swellhnd vagdry vagdis;
```

```
%LET n = 29 ;
%LET avarlist = ;
%DO i = 1 %TO &n ;
```

```

%LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
%LET avarlist = &avarlist a&var_i ;
%END;

DATA f38_ep_fu_pub (DROP = &avarlist );
SET f38_ep_fu_pub (RENAME = ( %DO i = 1 %TO &n ;
                            %LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
                            %LET avar_i = %SCAN(&avarlist, %EVAL(&i),%STR( ));
                            &var_i = &avar_i
                            %END;
                            ));

BY id year ;
ARRAY c(*) &avarlist ;
ARRAY d(*) &varlist ;
RETAIN &varlist ;

IF FIRST.id THEN DO;
DO i=1 TO DIM(c);
d(i)=c(i);
END;
END;
ELSE DO;
DO i=1 TO DIM(c);
IF c(i) NE . THEN d(i)=c(i);

END;
END;
IF LAST.year THEN OUTPUT;
RUN;

/* f45b_ep_fu_pub */
PROC SORT DATA=f45b_ep_fu_pub;
BY id year day &order vtyp;
RUN;

DATA f45b_ep_fu_pub;
SET f45b_ep_fu_pub;
BY id year day ;
IF FIRST.day ;
RUN;

%LET varlist = multi mvmin vite;
%LET n = 3 ;
%LET avarlist = ;
%DO i = 1 %TO &n;
%LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
%LET avarlist = &avarlist a&var_i;
%END;

DATA f45b_ep_fu_pub (DROP = &avarlist );
SET f45b_ep_fu_pub (RENAME = ( %DO i = 1 %TO &n ;
                            %LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
                            %LET avar_i = %SCAN(&avarlist, %EVAL(&i),%STR( ));
                            &var_i = &avar_i
                            %END;
                            ));

BY id year ;
ARRAY c(*) &avarlist ;
ARRAY d(*) &varlist ;
RETAIN &varlist ;

IF FIRST.id THEN DO;
DO i=1 TO DIM(c);
d(i)=c(i);
END;
END;
ELSE DO;
DO i=1 TO DIM(c);
IF c(i) NE . THEN d(i)=c(i);

END;
END;
IF LAST.year THEN OUTPUT;
RUN;

/* f80_ep_fu_pub */
PROC SORT DATA=f80_ep_fu_pub;
BY id year day &order vtyp;
RUN;

DATA f80_ep_fu_pub;
SET f80_ep_fu_pub;
BY id year day;
IF FIRST.day;
run;

%LET varlist = systol diastol bmi bmicx;
%LET n = 4 ;
%LET avarlist = ;
%DO i = 1 %TO &n ;
%LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));

```

```

%LET avarlist = &avarlist a&var_i ;
%END;

DATA f80_ep_fu_pub (DROP = &avarlist );
SET f80_ep_fu_pub (RENAME = ( %DO i = 1 %TO &n ;
                            %LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
                            %LET avar_i = %SCAN(&avarlist, %EVAL(&i),%STR( ));
                            &var_i = &avar_i
                            %END;
                            ));

BY id year ;
ARRAY c(*) &avarlist ;
ARRAY d(*) &varlist ;
RETAIN &varlist ;

IF FIRST.id THEN DO;
DO i=1 TO DIM(c);
d(i)=c(i);
END;
END;
ELSE DO;
DO i=1 TO DIM(c);
IF c(i) NE . THEN d(i)=c(i);

END;
END;
IF LAST.year THEN OUTPUT;
RUN;

/* f84_ep_fu_pub */
PROC SORT DATA=f84_ep_fu_pub;
BY id year day &order vtyp;
RUN;

DATA f84_ep_fu_pub;
SET f84_ep_fu_pub;
BY id year day;
IF FIRST.day;
RUN;

%LET varlist = brfinrt brfinlt;
%LET n = 2 ;
%LET avarlist = ;
%DO i = 1 %TO &n ;
%LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
%LET avarlist = &avarlist a&var_i ;
%END;

DATA f84_ep_fu_pub (DROP = &avarlist );
SET f84_ep_fu_pub (RENAME = ( %DO i = 1 %TO &n ;
                            %LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
                            %LET avar_i = %SCAN(&avarlist, %EVAL(&i),%STR( ));
                            &var_i = &avar_i
                            %END;
                            ));

BY id year ;
ARRAY c(*) &avarlist ;
ARRAY d(*) &varlist ;
RETAIN &varlist ;

IF FIRST.id THEN DO;
DO i=1 TO DIM(c);
d(i)=c(i);
END;
END;
ELSE DO;
DO i=1 TO DIM(c);
IF c(i) NE . THEN d(i)=c(i);

END;
END;
IF LAST.year THEN OUTPUT;
RUN;

/* f85_ep_fu_pub */
PROC SORT DATA=f85_ep_fu_pub;
BY id year day &order vtyp;
RUN;

DATA f85_ep_fu_pub;
SET f85_ep_fu_pub;
BY id year day;
IF FIRST.day;
RUN;

%LET varlist = rfureslt lfureslt;
%LET n = 2 ;
%LET avarlist = ;
%DO i = 1 %TO &n ;
%LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));

```



```

%LET avarlist = &avarlist a&var_i ;
%END;

DATA f85_ep_fu_pub (DROP = &avarlist );
SET f85_ep_fu_pub (RENAME = ( %DO i = 1 %TO &n ;
                             %LET var_i = %SCAN(&varlist, %EVAL(&i),%STR( ));
                             %LET avar_i = %SCAN(&avarlist, %EVAL(&i),%STR( ));
                             &var_i = &avar_i
                             %END;
                             ));

BY id year ;
ARRAY c(*) &avarlist ;
ARRAY d(*) &varlist ;
RETAIN &varlist ;

IF FIRST.id THEN DO;
  DO i=1 TO DIM(c);
    d(i)=c(i);
  END;
END;
ELSE DO;
  DO i=1 TO DIM(c);
    IF c(i) NE . THEN d(i)=c(i);
  END;
END;

IF LAST.year THEN OUTPUT;
RUN;

/*****
  MERGE FOLLOWUP DATA
  *****/
DATA followup;
MERGE f10_50
      f33
      f35_ep_fu_pub
      f38_ep_fu_pub
      f45b_ep_fu_pub
      f80_ep_fu_pub
      f84_ep_fu_pub
      f85_ep_fu_pub
      med_fu
      ht_adh_ep_fu_pub;
BY id year;
RUN;

PROC DATASETS LIBRARY = work NOLIST ;
DELETE f10_50 f33 f35_ep_fu_pub f38_ep_fu_pub f45b_ep_fu_pub f80_ep_fu_pub
       f84_ep_fu_pub f85_ep_fu_pub med_fu ht_adh_ep_fu_pub;
QUIT;

/*****
  MERGE BASELINE & FOLLOWUP DATA
  *****/
PROC SQL;
CREATE TABLE whi2 AS
SELECT M.*, S.*
FROM whi.baseline_bc M LEFT JOIN followup S
ON M.id=S.id;
QUIT;

/*-----
  IDENTIFY WHETHER A PATIENT MISSED A FOLLOWUP VISIT
  -----*/
/* create an artificial dataset with complete data up until maxyear */
PROC SQL;
CREATE TABLE aa AS
SELECT id, MAX(year) AS maxyear1
FROM followup
GROUP BY id;
QUIT;

PROC SQL;
CREATE TABLE bb AS
SELECT O.*, L.maxyear1
FROM whi.baseline_bc O, aa L
WHERE O.id=L.id;
QUIT;

DATA cc;
SET bb;
IF lastcont=. THEN DO;
  IF invady NE . AND deathdy NE . THEN lastcont=invady;
  ELSE IF invady NE . AND deathdy = . THEN lastcont=invady;
  ELSE IF invady = . AND deathdy NE . THEN lastcont=deathdy;
  ELSE IF invady = . AND deathdy = . and maxyear1 NE . THEN lastcont=maxyear1*365.25;
  ELSE lastcont=5.6*365.25;
END;

```

```

END;
maxyear=CEIL(lastcont/365.25);
DROP maxyear1;
RUN;

DATA dd;
SET cc;
year=0;
RETAIN year;
DO year=1 TO 9;
    OUTPUT;
END;
RUN;

/* maxyear copies of baseline covariates for merging into followup data */
DATA new;
SET dd;
IF year>maxyear THEN DELETE;
RUN;

/* add in time-varying variables to new from whi2 (left join will not overwrite common baseline variables) */
PROC SQL;
    CREATE TABLE whi AS
    SELECT O.*, L.*
    FROM new O LEFT JOIN whi2 L
    ON O.id=L.id AND O.year=L.year;
QUIT;

PROC SORT DATA=whi;
BY id year;
RUN;

/*-----
REPLACE MISSING VALUE IN A GIVEN VISIT WITH THE MOST RECENT VISIT W/O MISSING VALUE
-----*/

DATA whi_a;
SET whi;
BY id;
ARRAY e(*)
    /* f10_50_ep_fu_pub */
    hyst vagbleed heavybld vstartdy intermit bleednow vstopdy brsttend sevtend brstchn g corticos
    anticoag testost tamoxif endohyp hightrig trigl000 bldclot melanoma heartstk meningio brstcanc
    gallblad pancreas contmeds rtrneval gynnote physnote medchn g otheract pillfreq daysmiss allpills
    expsymp frgtpill frgtbotl tookbrk afraid famrcmnd mdrcmnd nopills othrmiss iap recontac
    /* f33_ep_fu_pub */
    f33faint f33falls physexam eyeexam brstexam mammogrm brstbpsy rctlexam hemoclt flexsig barium
    bpcheck cholchk ecg ptca outpdvt papsmr dandc endobpsy
    /* f35_ep_fu_pub */
    hrdex modex mldex smoking cigsdays tepiwk lepitot phyact texpwk alcswk
    /* f38_ep_fu_pub */
    lifequal genhel bodpain constip diarrhea tired skindry upstom feltdep incont physfun symptom forget
    trbsleep bloating nightswt aches brstten hotflash moodswng lowbackp neckpain headache vagitch concn
    jntpain swellhnd vagdry vagdis
    /* f45b_ep_fu_pub */
    multi mvmin vite
    /* f80_ep_fu_pub */
    systol diastol bmi bmicx
    /* f84_ep_fu_pub */
    brfinrt brfinlt
    /* f85_ep_fu_pub */
    rfureslt lfureslt
    /* med_fu */
    statin serm bispho hilipid hypertension diabetes corticosteroid anticoagulant aspirin tamoxifen testosterone;
ARRAY f(*)
    /* f10_50_ep_fu_pub */
    ahyst avagbleed aheavybld avstartdy aintermit ablednow avstopdy abrsttend asevtend abrstchn g acorticos
    aanticoag atestost atamoxif aendohyp ahightrig atrigl000 abldclot amelanoma aheartstk ameningio abrstcanc
    agallblad apancreas acontmeds artrneval agynnote aphysnote amedchn g aotheract apillfreq adaysmiss aallpills
    aexpsymp afrgtpill afrgtbotl atookbrk aafraid afamrcmnd amdrcmnd anopills aothrmiss aiap arecontac
    /* f33_ep_fu_pub */
    af33faint af33falls aphysexam aeyeexam abrstexam amammogrm abrstbpsy arctlexam ahemoclt aflexsig abarium
    abpcheck acholchk aecg aptca aoutpdvt apapsmr adandc aendobpsy
    /* f35_ep_fu_pub */
    ahrdex amodex amldex asmoking acigsdays atepiwk alepitot aphyact atexpwk aalcswk
    /* f38_ep_fu_pub */
    alifequal agenhel abodpain aconstip adiarrhea atired askindry aupstom afeltdep aincont aphysfun asymptom
    aforget atrbsleep abloating anightswt aaches abrstten ahotflash amoodswng alowbackp aneckpain aheadline
    avagitch aconcn ajntpain aswellhnd avagdry avagdis
    /* f45b_ep_fu_pub */
    amulti amvmin avite
    /* f80_ep_fu_pub */
    asystol adiaastol abmi abmicx
    /* f84_ep_fu_pub */
    abrfinrt abrfinlt
    /* f85_ep_fu_pub */
    arfureslt alfureslt
    /* med_fu */
    astatin aserm abispho ahilipid ahypertension adiabates acorticosteroid aanticoagulant aaspirin atamoxifen
    atestosterone;

```

```

RETAIN hyst avagbleed heavybld avstartdy avintermit ableednow avstopdy abrsttend asevtend abrstchn g corticos aanticog aaticog
atestost atamoxif endohyp ahightrig atrig1000 abldclot amelanoma aheartstk ameningio abrstcanc agallblad apancreas
acontmeds artrneval agynnote aphysnote amedchn g aotheract apillfreq adaysmiss aallpills aexpsymp afrgtpill afrgtbotl
atookbrk aafraid afamrcmnd amdrcmnd anopills aotrmiss aiap arecontac af33faint af33falls aphysexam aeyeexam abrstexam
amammogr m abrstbpsy arctlexam ahemoclt aflexsig abarium abpcheck acholchk aecg aptca aoutpdvt apapsm r adandc aendobpsy
ahrdex amodex amldex asmoking acigsday atepiwk alepitot aphyact atexpwk aalcswk alifequal agenhel abodpain aconstip
adiarrhea atired askindry aupstom afeltdep aincont aphysfun asymptom aforget atrbsleep abloating anightswt aaches
abrstten ahotflash amoodswng alowbackp aneckpain aheadache avagitch aconcen ajntpain aswellhnd avagdry avagdis
amulti amvmin avite asystol adiaistol abmi abmicx abrfinrt abrfinlt arfureslt alfureslt astatin aserm abispho
ahilipid ahypertension diabetes acorticosteroid aanticogulant aaspirin atamoxifen atestosterone;
IF FIRST.id THEN DO;
  DO i=1 TO DIM(e);
    f(i)=e(i);
  END;
END;
ELSE DO;
  DO i=1 TO DIM(e);
    IF e(i) NE . THEN f(i)=e(i);
  END;
END;
DROP hyst vagbleed heavybld vstartdy intermit bleednow vstopdy brsttend sevtend brstchn g corticos anticoag testost tamoxif
endohyp hightrig trig1000 bldclot melanoma heartstk meningio brstcanc gallblad pancreas contmeds rtrneval gynnote
physnote medchn g otheract pillfreq daysmiss allpills expsymp frgtpill frgtbotl tookbrk afraid famrcmnd mdrcmnd
nopills otrmiss iap recontac f33faint f33falls physexam eyeexam brstexam mammogr m brstbpsy rctlexam hemoclt flexsig
barium bpcheck cholchk ecg ptca outpdvt papsm r dandc endobpsy hrdex modex mldex smoking cigday tepiwk lepitot
phyact texpwk alcswk lifequal genhel bodpain constip diarrhea tired skindry upstom feltdep incont physfun symptom
forget trbsleep bloating nightswt aches brstten hotflash moodswng lowbackp neckpain headache vagitch concen jntpain
swellhnd vagdry vagdis multi mvmin vite systol diastol bmi bmicx brfinrt brfinlt rfureslt lfureslt statin serm bispho
hilipid hypertension diabetes corticosteroid anticoagulant aspirin tamoxifen testosterone i;
RUN;

```

```

PROC DATASETS LIBRARY=work NOLIST ;
MODIFY whi_a ;
RENAME ahyst=hyst avagbleed=vagbleed aheavybld=heavybld avstartdy=vstartdy avintermit=intermit ableednow=bleednow
avstopdy=vstopdy abrsttend=brsttend asevtend=sevtend abrstchn g=brstchn g acorticos=corticos aanticog=anticog
atestost=testost atamoxif=tamoxif endohyp=endohyp ahightrig=hightrig atrig1000=trig1000 abldclot=bldclot
amelanoma=melanoma aheartstk=heartstk ameningio=meningio abrstcanc=brstcanc agallblad=gallblad apancreas=pancreas
acontmeds=contmeds artrneval=rtrneval agynnote=gynnote aphysnote=physnote amedchn g=medchn g aotheract=otheract
apillfreq=pillfreq adaysmiss=daysmiss aallpills=allpills aexpsymp=expsymp afrgtpill=frgtpill afrgtbotl=frgtbotl
atookbrk=tookbrk aafraid=afraid afamrcmnd=famrcmnd amdrcmnd=mdrcmnd anopills=nopills aotrmiss=otrmiss aiap=iap
arecontac=recontac af33faint=f33faint af33falls=f33falls aphysexam=physexam aeyeexam=eyeexam abrstexam=brstexam
amammogr m=mammogr m abrstbpsy=brstbpsy arctlexam=rctlexam ahemoclt=hemoclt aflexsig=flexsig abarium=barium
abpcheck=bpcheck acholchk=cholchk aecg=ecg aptca=ptca aoutpdvt=outpdvt apapsm r=papsm r adandc=dandc aendobpsy=endobpsy
ahrdex=hrdex amodex=modex amldex=mldex atepiwk=tepiwk alepitot=lepitot asmoking=smoking acigsday=cigsday
aphyact=phyact atexpwk=texpwk aalcswk=alcswk alifequal=lifequal agenhel=genhel abodpain=bodpain aconstip=constip
adiarrhea=diarrhea atired=tired askindry=skindry aupstom=upstom afeltdep=feltdep aincont=incont aphysfun=physfun
asymptom=symptom aforget=forget atrbsleep=trbsleep abloating=bloating anightswt=nightswt aaches=aches abrstten=brstten
ahotflash=hotflash amoodswng=moodswng alowbackp=lowbackp aneckpain=neckpain aheadache=headache avagitch=vagitch
aconcen=concen ajntpain=jntpain aswellhnd=swellhnd avagdry=vagdry avagdis=vagdis amulti=multi amvmin=mvmin avite=vite
asystol=systol adiaistol=diastol abmi=bmi abmicx=bmicx abrfinrt=brfinrt abrfinlt=brfinlt arfureslt=rfureslt
alfureslt=lfureslt astatin=statin aserm=serm abispho=bispho ahilipid=hilipid ahypertension=hypertension
diabetes=diabetes acorticosteroid=corticosteroid aanticogulant=anticogulant aaspirin=aspirin atamoxifen=tamoxifen
atestosterone=testosterone;
QUIT;

```

```

/*-----
RECODE VARIABLES (BASELINE & FOLLOWUP)
-----*/
DATA whi_a;
SET whi_a;
/* f10_50_ep_fu_pub */
ARRAY a(*) hyst vagbleed heavybld vstartdy intermit bleednow vstopdy brsttend sevtend brstchn g corticos anticoag testost
tamoxif endohyp hightrig trig1000 bldclot melanoma heartstk meningio brstcanc gallblad pancreas contmeds
rtrneval gynnote physnote otheract pillfreq daysmiss allpills expsymp frgtpill frgtbotl tookbrk
afraid famrcmnd mdrcmnd nopills otrmiss iap recontac;
DO b=1 TO DIM(a);
  IF a(b)=. THEN a(b)=0;
END;

/* f33_ep_fu_pub */
ARRAY c(*) f33faint f33falls physexam eyeexam brstexam mammogr m brstbpsy rctlexam hemoclt flexsig barium
bpcheck cholchk ecg ptca outpdvt papsm r dandc endobpsy ;
DO d=1 TO DIM(c);
  IF c(d)=. THEN c(d)=0;
END;

/* f35_ep_fu_pub */;
IF texpwk= . THEN met=9;
IF texpwk=0 THEN met=0;
IF 0.00<texpwk<= 4.25 THEN met=1;
IF 4.25<texpwk<=10.00 THEN met=2;
IF 10.00<texpwk<=20.00 THEN met=3;
IF texpwk> 20.00 THEN met=4;
IF smoking=. THEN smoking=9;

IF alcswk = 0 AND alcoholb=1 THEN alcohol=1;
IF alcswk = 0 AND 1<alcoholb<9 THEN alcohol=2;
IF alcswk = 0 AND alcoholb=9 THEN alcohol=1;
IF alcswk = . THEN alcohol=9;
IF 0 < alcswk <= 0.25 THEN alcohol=3;

```

```

IF 0.25 < alcswk <= 1      THEN alcohol=4;
IF 1 < alcswk < 7         THEN alcohol=5;
IF alcswk >= 7            THEN alcohol=6;

/* f38_ep_fu_pub */
IF genhel=.                THEN genhel=9;
IF bodpain=.              THEN bodpain=9;
IF physfun=.              THEN physfun=0;
IF physfun<=75            THEN physfun=1;
IF 75<physfun<=90         THEN physfun=2;
IF 90<physfun<=95         THEN physfun=3;
IF physfun> 95            THEN physfun=4;

IF constip in (.,0)       THEN constip=0; ELSE constip=1;
IF diarrhea in (.,0)      THEN diarrhea=0; ELSE diarrhea=1;
IF upstom in (.,0)        THEN upstom=0; ELSE upstom=1;
IF feltdep in (.,0)       THEN feltdep=0; ELSE feltdep=1;
IF incont=.               THEN incont=0;
IF skindry in (.,0)       THEN skindry=0;
IF tired in (.,0)         THEN tired=0;
IF forget in (.,0)        THEN forget=0;
IF trbsleep in (.,0)      THEN trbsleep=0;
IF symptom=.              THEN symptom=0;

IF bloating in (.,0)      THEN bloating=0;
IF nightswt in (.,0)      THEN nightswt=0;
IF aches in (.,0)         THEN aches=0;
IF brstten in (.,0)       THEN brstten=0;
IF hotflash in (.,0)      THEN hotflash=0;
IF moodswng in (.,0)      THEN moodswng=0;
IF lowbackp in (.,0)      THEN lowbackp=0;
IF neckpain in (.,0)      THEN neckpain=0;
IF headache in (.,0)      THEN headache=0;
IF vagitch in (.,0)       THEN vagitch=0;
IF concen in (.,0)        THEN concen=0;
IF jntpain in (.,0)       THEN jntpain=0;
IF swellhnd in (.,0)      THEN swellhnd=0;
IF vagdry in (.,0)        THEN vagdry=0;
IF vagdis in (.,0)        THEN vagdis=0;

/* f45b_ep_base_pub */
IF multi=1 OR mvmin=1     THEN multi=1;
ELSE multi=0;
IF vite=1                  THEN vite=1;
ELSE vite=0;

/* f80_ep_fu_pub */
IF systol=.                THEN systol=1;
IF diastol=.               THEN diastol=1;
IF bmicx=.                 THEN bmicx=9;

/* f84_ep_fu_pub */
IF brfinrt= . OR brfinlt= . THEN brfin=9;
IF brfinrt= 0 AND brfinlt= 0 THEN brfin=0;
IF brfinrt= 1 AND brfinlt<=1 THEN brfin=1;
IF brfinrt<=1 AND brfinlt= 1 THEN brfin=1;
IF brfinrt> 1 OR brfinlt> 1 THEN brfin=2;

/* f85_ep_fu_pub */
IF rfureslt= . OR lfureslt= . THEN mafin=9;
IF rfureslt= 0 AND lfureslt= 0 THEN mafin=0;
IF rfureslt= 1 AND lfureslt<=1 THEN mafin=1;
IF rfureslt<=1 AND lfureslt= 1 THEN mafin=1;
IF rfureslt> 1 OR lfureslt> 1 THEN mafin=2;

IF brfin in (0,9) AND mafin in (0,9) THEN brmafin=0;
IF brfin = 1 AND mafin in (0,1,9) THEN brmafin=1;
IF brfin in (0,1,9) AND mafin = 1 THEN brmafin=1;
IF brfin = 2 OR mafin = 2 THEN brmafin=2;
IF brfin = 9 AND mafin = 9 THEN brmafin=9;

/* med_fu */
IF statin=.                THEN statin=0;
IF serm=.                  THEN serm=0;
IF bispho=.                THEN bispho=0;
IF hilipid=.               THEN hilipid=0;
IF hypertension=.          THEN hypertension=0;
IF diabetes=.              THEN diabetes=0;
IF corticosteroid=.        THEN corticosteroid=0;
IF anticoagulant=.         THEN anticoagulant=0;
IF aspirin=.               THEN aspirin=0;
IF tamoxifen=.              THEN tamoxifen=0;
IF testosterone=.          THEN testosterone=0;
IF estpro=.                THEN estpro=0;
IF serm=0 AND (tamoxif=1 OR tamoxifen=1) THEN serm=1;
DROP b d;

IF (diabslf=1 & year>=diabslfdy/365.25) OR diabetes=1 THEN diabslf=1;
ELSE diabslf=0;
IF (hyppill=1 & year>=hyppildy/365.25) OR (systol=3 & diastol=2) OR hypertension=1 THEN htn=1;
ELSE htn=0;

```

```

IF brstc=1 & year>=brstcdy/365.25 THEN breast=1;
ELSE breast=0;
IF colotot=1 & year>=colototdy/365.25 THEN colotot=1;
ELSE colotot=0;
IF endoca=1 & year>=endocady/365.25 THEN endoca=1;
ELSE endoca=0;
IF ovca=1 & year>=ovcady/365.25 THEN ovca=1;
ELSE ovca=0;
IF bladder=1 & year>=bladderdy/365.25 THEN bladder=1;
ELSE bladder=0;
IF kidney=1 & year>=kidneydy/365.25 THEN kidney=1;
ELSE kidney=0;
IF leuk=1 & year>=leukdy/365.25 THEN leuk=1;
ELSE leuk=0;
IF lung=1 & year>=lungdy/365.25 THEN lung=1;
ELSE lung=0;
IF lymphnh=1 & year>=lymphnhdy/365.25 THEN lymphnh=1;
ELSE lymphnh=0;
IF mel=1 & year>=meldy/365.25 THEN mel=1;
ELSE mel=0;
IF mult=1 & year>=multdy/365.25 THEN mult=1;
ELSE mult=0;
IF panc=1 & year>=pancdy/365.25 THEN panc=1;
ELSE panc=0;
IF thy=1 & year>=thydy/365.25 THEN thy=1;
ELSE thy=0;
IF lothca=1 & year>=lothcady/365.25 THEN lothca=1;
ELSE lothca=0;
IF acs=1 & year>=acsdly/365.25 THEN acs=1;
ELSE acs=0;
IF chf=1 & year>=chfdy/365.25 THEN chf=1;
ELSE chf=0;
IF revasc=1 & year>=revascdy/365.25 THEN revasc=1;
ELSE revasc=0;
IF pvd=1 & year>=pvddy/365.25 THEN pvd=1;
ELSE pvd=0;
IF dvt=1 & year>=dvtidy/365.25 THEN dvt=1;
ELSE dvt=0;
IF pe=1 & year>=pedy/365.25 THEN pe=1;
ELSE pe=0;
IF vte=1 & year>=vtedy/365.25 THEN vte=1;
ELSE vte=0;
IF stroke=1 & year>=strokeddy/365.25 THEN stroke=1;
ELSE stroke=0;
IF fracadj=1 & year>=fracdy/365.25 THEN fracadj=1;
ELSE fracadj=0;
IF osteop=1 & year>=osteopdy/365.25 THEN osteop=1;
ELSE osteop=0;
IF osteoart=1 & year>=ostartdy/365.25 THEN osteoart=1;
ELSE osteoart=0;
IF rheumart=1 & year>=rheumdy/365.25 THEN rheumart=1;
ELSE rheumart=0;
IF heart=1 & year>=heartdy/365.25 THEN heart=1;
ELSE heart=0;
IF cathhosp=1 & year>=cthslyfdy/365.25 THEN cathhosp=1;
ELSE cathhosp=0;
IF othheart=1 & year>=othhrtdy/365.25 THEN othheart=1;
ELSE othheart=0;
IF cataract=1 & year>=catardy/365.25 THEN cataract=1;
ELSE cataract=0;
IF glaucoma=1 & year>=glaucdy/365.25 THEN glaucoma=1;
ELSE glaucoma=0;
IF lupus=1 & year>=lupusdy/365.25 THEN lupus=1;
ELSE lupus=0;
IF intpoly=1 & year>=intplydy/365.25 THEN intpoly=1;
ELSE intpoly=0;
IF kidstone=1 & year>=kidstndy/365.25 THEN kidstone=1;
ELSE kidstone=0;
IF choleitis=1 & year>=choleddy/365.25 THEN choleitis=1;
ELSE choleitis=0;
IF stone=1 & year>=stonedy/365.25 THEN stone=1;
ELSE stone=0;
IF gallproc=1 & year>=gallpdy/365.25 THEN gallproc=1;
ELSE gallproc=0;
IF ooph=1 & year>=oophdy/365.25 THEN ooph=1;
ELSE ooph=0;
IF hospever=1 & year>=hospdy/365.25 THEN hospever=1;
ELSE hospever=0;

```

```

IF invasive=1 THEN DO; /* events and bc occurred during the same year, but bc occurred first */

```

```

IF diabslf=1 & int(invady/365.25)=int(diabslfdy/365.25) & invady<diabslfdy THEN diabslf=0;
IF htn=1 & int(invady/365.25)=int(hyppildy/365.25) & invady<hyppildy THEN htn=0;
IF chd=1 & int(invady/365.25)=int(chddy/365.25) & invady<chddy THEN chd=0;
IF colotot=1 & int(invady/365.25)=int(colototdy/365.25) & invady<colototdy THEN colotot=0;
IF endoca=1 & int(invady/365.25)=int(endocady/365.25) & invady<endocady THEN endoca=0;
IF ovca=1 & int(invady/365.25)=int(ovcady/365.25) & invady<ovcady THEN ovca=0;
IF bladder=1 & int(invady/365.25)=int(bladderdy/365.25) & invady<bladderdy THEN bladder=0;
IF kidney=1 & int(invady/365.25)=int(kidneydy/365.25) & invady<kidneydy THEN kidney=0;
IF leuk=1 & int(invady/365.25)=int(leukdy/365.25) & invady<leukdy THEN leuk=0;
IF lung=1 & int(invady/365.25)=int(lungdy/365.25) & invady<lungdy THEN lung=0;
IF lymphnh=1 & int(invady/365.25)=int(lymphnhdy/365.25) & invady<lymphnhdy THEN lymphnh=0;

```

```

IF mel=1 & int(invady/365.25)=int(meldy/365.25) & invady<meldy THEN mel=0;
IF mult=1 & int(invady/365.25)=int(multdy/365.25) & invady<multdy THEN mult=0;
IF panc=1 & int(invady/365.25)=int(pancdy/365.25) & invady<pancdy THEN panc=0;
IF thy=1 & int(invady/365.25)=int(thydy/365.25) & invady<thydy THEN thy=0;
IF lothca=1 & int(invady/365.25)=int(lothcady/365.25) & invady<lothcady THEN lothca=0;
IF acs=1 & int(invady/365.25)=int(acsdly/365.25) & invady<acsdly THEN acs=0;
IF chf=1 & int(invady/365.25)=int(chfdy/365.25) & invady<chfdy THEN chf=0;
IF revasc=1 & int(invady/365.25)=int(revascdy/365.25) & invady<revascdy THEN revasc=0;
IF pvd=1 & int(invady/365.25)=int(pvddy/365.25) & invady<pvddy THEN pvd=0;
IF dvt=1 & int(invady/365.25)=int(dvtdy/365.25) & invady<dvtdy THEN dvt=0;
IF pe=1 & int(invady/365.25)=int(pedy/365.25) & invady<pedy THEN pe=0;
IF vte=1 & int(invady/365.25)=int(vtedy/365.25) & invady<vtedy THEN vte=0;
IF stroke=1 & int(invady/365.25)=int(strokeddy/365.25) & invady<strokeddy THEN stroke=0;
IF fracadj=1 & int(invady/365.25)=int(fracdy/365.25) & invady<fracdy THEN fracadj=0;
IF osteop=1 & int(invady/365.25)=int(osteopdy/365.25) & invady<osteopdy THEN osteop=0;
IF osteoart=1 & int(invady/365.25)=int(ostartdy/365.25) & invady<ostartdy THEN osteoart=0;
IF rheumart=1 & int(invady/365.25)=int(rheumdy/365.25) & invady<rheumdy THEN rheumart=0;
IF heart=1 & int(invady/365.25)=int(heartdy/365.25) & invady<heartdy THEN heart=0;
IF cathhosp=1 & int(invady/365.25)=int(cthslfdy/365.25) & invady<cthslfdy THEN cathhosp=0;
IF othheart=1 & int(invady/365.25)=int(othhrtdy/365.25) & invady<othhrtdy THEN othheart=0;
IF cataract=1 & int(invady/365.25)=int(catarady/365.25) & invady<atarady THEN cataract=0;
IF glaucoma=1 & int(invady/365.25)=int(glaucdy/365.25) & invady<glaucdy THEN glaucoma=0;
IF lupus=1 & int(invady/365.25)=int(lupusdy/365.25) & invady<lupusdy THEN lupus=0;
IF intpoly=1 & int(invady/365.25)=int(intplydy/365.25) & invady<intplydy THEN intpoly=0;
IF kidstone=1 & int(invady/365.25)=int(kidstndy/365.25) & invady<kidstndy THEN kidstone=0;
IF choleitis=1 & int(invady/365.25)=int(choleddy/365.25) & invady<choleddy THEN choleitis=0;
IF stone=1 & int(invady/365.25)=int(stoneddy/365.25) & invady<stoneddy THEN stone=0;
IF gallproc=1 & int(invady/365.25)=int(gallpdy/365.25) & invady<gallpdy THEN gallproc=0;
IF ooph=1 & int(invady/365.25)=int(oophdy/365.25) & invady<oophdy THEN ooph=0;
IF hospever=1 & int(invady/365.25)=int(hospdy/365.25) & invady<hospdy THEN hospever=0;
end;

```

```

menopause=SUM(bloating,nightswt,aches,brstten,hotflash,moodswng,lowbackp,neckpain,headache,vagitch,concen,
jntpain,swellhnd,vagdry,vagdis);

```

```

IF 0<=menopause<=3 THEN menosym=0;
IF 3< menopause<=6 THEN menosym=1;
IF 6< menopause<=10 THEN menosym=2;
IF menopause> 10 THEN menosym=3;

```

```

IF hotflash=0 AND nightswt=0 THEN vasomotor=0;
ELSE IF hotflash=0 AND nightswt=1 THEN vasomotor=1;
ELSE IF hotflash=1 AND nightswt=0 THEN vasomotor=1;
ELSE IF hotflash=1 AND nightswt=1 THEN vasomotor=1;
ELSE vasomotor=2;

```

```

IF diabtrtb=1 THEN diabslf=1;
ELSE diabslf=diabslf;
IF htnb=1 THEN htn=1;
ELSE htn=htn;
IF hicholb=1 OR hilipid=1 THEN hichol=1;
ELSE hichol=0;

```

```

cardio=SUM(acs,chf,revasc,pvd,dvt,pe,stroke,heart,cathhosp,othheart);
IF cardio>=3 THEN cardio=3;

```

```

cancer=(SUM(breast,colotot,endoca,ovca,bladder,kidney,leuk,lung,lymphnh,mel,mult,panc,thy,lothca)>0);

```

```

prevent=SUM(physexam,eyeexam,brstexam,mammogrm,brstbpsy,rctlexam,hemoclt,flexsig,barium,bpcheck,ecg,papsmr,
dandc,endobpsy,cholchk);
IF prevent>=6 THEN prevent=6;

```

```

/*-----
CHANGE IN BMI, PHYSICAL ACTIVITY & ALCOHOL INTAKE FROM BASELINE
-----*/

```

```

cbmi = bmi - bmib;
cmet = texpwk - texpwkb;
calc = alcswk - alcswkb;

```

```

/*-----
EVENTS THAT LEAD TO PERMANENT STOPPING OF STUDY HRT
-----*/

```

```

IF dvt=1 OR /* deep vein thrombosis */
pe=1 OR /* pulmonary embolus */
anticoag=1 OR /* anticoagulants */
anticoagulant=1 OR /* anticoagulants (drug file) */
endohyp=1 OR /* endometrial hyperplasia */
trig1000=1 OR /* triglycerides above 1000mg/dl */
estrogen=1 OR /* estrogen */
progest=1 OR /* progesterone */
estpro=1 OR /* estrogen + progesterone */
testost=1 OR /* testosterone */
testosterone=1 OR /* testosterone (drug file) */
tamoxif=1 OR /* tamoxifen */
tamoxifen=1 OR /* tamoxifen (drug file) */
serm=1 OR /* serms */
melanoma=1 OR /* malignant melanoma */
endoca=1 OR /* endometrial cancer */
THEN stop1=1; ELSE stop1=0;

```

```

IF estrogen=1 OR
  progest=1 OR
  estpro=1
THEN stop2=1;
ELSE stop2=0;

```

```

DROP brfinrt brfinlt rfureslt lfureslt;
RUN;

```

```

/*-----

```

```

LAG THE DATA
-----*/

```

```

PROC SORT DATA=whi_a;
BY id year;
RUN;

```

```

DATA whi_a;
SET whi_a;
BY id;
ARRAY a(*) chd cardio cancer endohyp hyst ooph hightrig hichol statin aspirin serm bispho multi vite
  diabslf htn smoking met bmicx alcohol cbmi cmet calc vagbleed heavybld intermit bleednow
  brsttend sevtend brstchng fracadj osteop prevent brstexam mammogrm brstbpsy brfin mafin
  brmafin hospever genhel bodpain feltdep physfun symptom menosym hotflash nightswt vasomotor;

```

```

ARRAY b(*) lchd lcardio lcancer lendohyp lhyst looph lhightrig lhichol lstatin laspirin lserm lbispho lmulti
  lvite ldiabslf lhtn lsmoking lmet lbmicx lalcohol lcbmi lcmet lcalc lvagbleed lheavybld
  lintermit lbleednow lbrsttend lsevtend lbrstchng lfracadj losteop lprevent lbrstexam
  lmammogrm lbrstbpsy lbrfin lmafin lbrmafin lhospesever lgenhel lbodpain lfeltdep lphysfun
  lsymptom lmenosym lhotflash lnightswt lvasomotor;

```

```

DO i=1 TO DIM(a);
  b(i)=LAG1(a(i));
END;

```

```

IF FIRST.id THEN DO;
  lchd      = chdb;
  lcardio   = SUM(dvtb,peb,strokeb,padb,cvdb,arrestb,chfb,cathb,carotidb,afb,aneurysmb,tiab,anginab);
  IF lcardio>=3 THEN lcardio=3;
  lcancer   = cancerb;
  lendohyp  = 0;
  lhyst     = hystb;
  looph     = boophb;
  lhightrig = 0;
  lhichol   = hicholb;
  lstatin   = statinb;
  laspirin  = aspirinb;
  lserm     = sermb;
  lbispho   = bisphob;
  lmulti    = multib;
  lvite     = viteb;
  ldiabslf  = diabtrtb;
  lhtn      = htnb;
  lsmoking  = smokingb;
  IF smokingb IN (0,9) THEN lsmoking=smokingb;
  IF smokingb=1 THEN lsmoking=0;
  IF smokingb=2 THEN lsmoking=1;
  lmet      = metb;
  lbmicx    = bmicxb;
  lalcohol  = alcoholb;
  lcbmi     = 0;
  lcmet     = 0;
  lcalc     = 0;
  lvagbleed = 0;
  lheavybld = 0;
  lintermit = 0;
  lbleednow = 0;
  lbrsttend = brsttendb;
  lsevtend  = sevtendb;
  lbrstchng = 0;
  lfracadj  = bkboneb;
  losteop   = osteopb;
  lbrstexam = 1;
  lmammogrm = mammob;
  lbrstbpsy = brstbiopb;
  lbrfin    = brfinb;
  lmafin    = mafinb;
  lbrmafin  = brmafinb;
  lhospesever = hosp2y;
  lprevent  = SUM(mammob,papsmearb,brstbiopb,endoasp,colnscpyb,pcolonrmb,hemocculb,nedlaspb,boophb);
  IF lprevent>=6 THEN lprevent=6;
  lgenhel   = genhelb;
  lbodpain  = bodpainb;
  lfeltdep  = feltdepb;
  lphysfun  = physfunb;
  lsymptom  = symptomb;
  lmenosym  = menosymb;
  lhotflash = hotflashb;
  lnightswt = nightswtb;
  lvasomotor = vasomotorb;
END;

```

```
DROP i;
RUN;

/*****
  CREATE A PERMANENT SAS DATASET FOR IPW ANALYSIS
  *****/
DATA whi.ipwdr_bc;
SET whi_a;
IF invasive=1 AND year > invayear THEN DELETE;
IF invasive=1 AND year = invayear THEN invasive=1;
ELSE invasive=0;
IF death=1 AND year > dyear THEN DELETE;
IF death=1 AND year = dyear THEN death=1;
ELSE death=0;
RUN;

PROC DATASETS LIBRARY=work NOLIST ;
DELETE a aa adh_a adh_c b bb cc dd f10_50_ep_fu_pub f10_50_ep_fu_publ f33_ep_fu_pub f44ref_meds
  f44_ct_pub f44_ct_publ f44_ep_base_pub f44_ep_fu_pub followup id med_ep_base med_ep_fu med_ep_ful new
  ;
QUIT;

%MEND data_management;
```

```

/*****
  IDENTIFY LOCATIONS OF SAS LIBRARY AND RAW DATASETS, CALL OUT MACROS
  *****/
LIBNAME whi 'c:\whi';

%LET order= ;

%LET file_location    = c:\whi\;
%LET file_location_fu = c:\whi\;

%data_management;
```





```

-----*/
DATA ipw_a; SET whi.ipwdr_bc (KEEP=
id r invasive invady invayear death deathdy dyear lastcont /* outcome related */
adhrate estrogen progest estpro openlabel pillfreq1 pillfreq2 /* hormone use */
year dmflag /* year and DM status */
region ethnic educ marital /* demographic factors */
menarche gravid parity brstfed brstdisb agefbir /* reproductive factors */
bcnum brcafrel bkboneb /* family history */
age ager diabtrtb smokingb metb bmicxb alcoholb /* established risk factors */
genhelb bodpainb physfunb preventb hopefulb /* physical & mental health */
sermb bisphob multib fruitb vegeb /* medication & diet */
chdb cardiob cancerb bkboneb osteopb /* cvd, cancer, fracture */
boophb nomam2yr /* mammogram & breast exam */
menoage menosymb vasomotorb /* menopause-related */
recency tothcat oc /* hrt and oc use */

```

```

ldiabslf lsmoking lmet lbmicx lalcohol /* established risk factors */
lgenhel lbodpain lphysfun lprevent /* physical & mental health */
lserm lbispho lmulti /* medication & diet */
lchd lcardio lcancer lfracadj losteop /* cvd, cancer, fracture */
looph lbrstexam lmammogrm lbrstbpsy lbrfin lmafin lbrmafin /* mammogram & breast exam */
lmenosym lvasomotor /* menopause-related */
lvagbleed lheavybld lintermit lbleednow lbrsttend lsevtend lbrstchnng; /* side effects */

```

```

/*-----
RECODE VARIABLES
-----*/

```

```

region2=(region=2);      region3=(region=3);      region4=(region=4);
ethnic3=(ethnic=3);      ethnic4=(ethnic=4);      ethnic8=(ethnic=8);
educ1=(educ=1);          educ2=(educ=2);          educ9=(educ=9);
marital1=(marital=1);    marital2=(marital=2);    marital3=(marital=3);
ageg2=(55<=age<60);      ageg3=(60<=age<65);      ageg4=(65<=age<70);      ageg5=(70<=age<75);
ageg6=(75<=age<80);
smokingb1=(smokingb=1);  smokingb2=(smokingb=2);  smokingb9=(smokingb=9);
metb1=(metb=1);          metb2=(metb=2);          metb3=(metb=3);          metb4=(metb=4);
metb9=(metb=9);
bmicxb3=(bmicxb=3);      bmicxb4=(bmicxb=4);      bmicxb5=(bmicxb=5);      bmicxb6=(bmicxb=6);
alcoholb2=(alcoholb=2);  alcoholb3=(alcoholb=3);  alcoholb4=(alcoholb=4);  alcoholb5=(alcoholb=5);
alcoholb6=(alcoholb=6);  alcoholb9=(alcoholb=9);
genhelb2=(genhelb=2);    genhelb3=(genhelb=3);    genhelb4=(genhelb=4);    genhelb5=(genhelb=5);
bodpainb2=(bodpainb=2);  bodpainb3=(bodpainb=3);  bodpainb4=(bodpainb=4);  bodpainb5=(bodpainb=5);
bodpainb9=(bodpainb=9);
physfunb2=(physfunb=2);  physfunb3=(physfunb=3);  physfunb4=(physfunb=4);
preventb1=(preventb=1);  preventb2=(preventb=2);  hopefulb3=(hopefulb=3);  hopefulb5=(hopefulb=5);
hopefulb1=(hopefulb=1);  hopefulb2=(hopefulb=2);
hopefulb9=(hopefulb=9);
fruitb1=(fruitb=1);      fruitb2=(fruitb=2);      fruitb3=(fruitb=3);
vegeb1=(vegeb=1);        vegeb2=(vegeb=2);        vegeb3=(vegeb=3);
menoage1=(menoage=1);    menoage2=(menoage=2);    menoage9=(menoage=9);
menosymb1=(menosymb=1);  menosymb2=(menosymb=2);  menosymb3=(menosymb=3);
recency1=(recency=1);    recency2=(recency=2);    recency3=(recency=3);    recency4=(recency=4);
tothcat1=(tothcat=1);    tothcat2=(tothcat=2);    tothcat3=(tothcat>=3);

lsmoking1=(lsmoking=1);  lsmoking9=(lsmoking=9);
lmet1=(lmet=1);          lmet2=(lmet=2);          lmet3=(lmet=3);          lmet4=(lmet=4);
lmet9=(lmet=9);
lbmicx1=(lbmicx=1);      lbmicx3=(lbmicx=3);      lbmicx4=(lbmicx=4);      lbmicx5=(lbmicx=5);
lbmicx6=(lbmicx=6);      lbmicx9=(lbmicx=9);
lalcohol2=(lalcohol=2);  lalcohol3=(lalcohol=3);  lalcohol4=(lalcohol=4);  lalcohol5=(lalcohol=5);
lalcohol6=(lalcohol=6);  lalcohol9=(lalcohol=9);
lgenhel2=(lgenhel=2);    lgenhel3=(lgenhel=3);    lgenhel4=(lgenhel=4);    lgenhel5=(lgenhel=5);
lgenhel9=(lgenhel=9);
lbodpain2=(lbodpain=2);  lbodpain3=(lbodpain=3);  lbodpain4=(lbodpain=4);  lbodpain5=(lbodpain=5);
lbodpain9=(lbodpain=9);
lphysfun2=(lphysfun=2);  lphysfun3=(lphysfun=3);  lphysfun4=(lphysfun=4);
lprevent1=(lprevent=1);  lprevent2=(lprevent=2);  lprevent3=(lprevent=3);  lprevent4=(lprevent=4);
lprevent5=(lprevent=5);  lprevent6=(lprevent=6);
lcardio1=(lcardio=1);    lcardio2=(lcardio=2);    lcardio3=(lcardio=3);
lmenosym1=(lmenosym=1);  lmenosym2=(lmenosym=2);  lmenosym3=(lmenosym=3);
lheavybld1=(lheavybld=1); lheavybld2=(lheavybld=2); lheavybld3=(lheavybld=3); lheavybld4=(lheavybld=4);
lsevtend1=(lsevtend=1);  lsevtend2=(lsevtend>=2);

menarche1=(menarche=1);  menarche2=(menarche=2);
gravid1=(gravid=1);      gravid2=(gravid=2);      gravid3=(gravid=3);      gravid9=(gravid=9);
parity1=(parity=1);      parity2=(parity in (2,3)); parity3=(parity in (4,5)); parity4=(parity>5);
parity9=(parity=9);
brstfed1=(brstfed=1);    brstfed2=(brstfed=2);
brstdisb1=(brstdisb=1);  brstdisb2=(brstdisb=2);  brstdisb9=(brstdisb=9);
agefbir1=(agefbir=1);    agefbir2=(agefbir=2);    agefbir3=(agefbir=3);    agefbir9=(agefbir=9);
vasomotorb1=(vasomotorb=1); vasomotorb2=(vasomotorb>1);

lbrfin0=(lbrfin=0);      lbrfin1=(lbrfin=1);      lbrfin2=(lbrfin=2);
lmafin0=(lmafin=0);      lmafin1=(lmafin=1);      lmafin2=(lmafin=2);
lbrmafin0=(lbrmafin=0);  lbrmafin1=(lbrmafin=1);  lbrmafin2=(lbrmafin=2);
lvasomotor1=(lvasomotor=1); lvasomotor2=(lvasomotor>1);

```

```

DROP smokingb metb bmicxb region ethnic educ marital hopefulb genhelb bodpainb fruitb vegeb preventb menosymb
physfunb alcoholb
lsmoking lmet lbmicx lgenhel lbodpain lcardio lphysfun lalcohol lmenosym lheavybld lsevtend lprevent

```

menarche parity brstfed brstdisb lbrfin lmafin lbrmafin; RUN;

```
/*-----  
  ASSIGN ADHRATE FOR THOSE W/ MISSING VALUES  
-----*/  
PROC SORT TAGSORT DATA=ipw_a; BY id year; RUN;  
  
DATA ipw0  _idtmp_ (keep = newid id );  
SET ipw_a (RENAME=(adhrate=adhrate_old)); BY id year;  
  
  retain newid ;  
  if _n_ = 1 then newid = 0 ;  
  if first.id then do ;  
    newid = newid + 1 ;  
    output _idtmp_ ;  
  end;  
  
adhrate = adhrate_old;  
d      = RANUNI(1232);  
  
IF adhrate_old=. THEN DO;  
  IF pillfreq1=0 THEN DO;  
    IF pillfreq2=0 THEN adhrate = (0.0 + 0.0) / 14;          /* not at all          */  
    IF pillfreq2=1 THEN adhrate = (0.0 + 0.5) / 14;          /* < 1 day per week   */  
    IF pillfreq2=2 THEN adhrate = (0.0 + 1.5) / 14;          /* 1-2 days per week  */  
    IF pillfreq2=3 THEN adhrate = (0.0 + 3.5) / 14;          /* 3-4 days per week  */  
    IF pillfreq2=4 THEN adhrate = (0.0 + 5.5) / 14;          /* 5-6 days per week  */  
    IF pillfreq2=5 THEN adhrate = (0.0 + 7.0) / 14;          /* every day of the week */  
    IF pillfreq2=. THEN adhrate = (0.0 + 7*d) / 14;          /* missing             */  
  END;  
  IF pillfreq1=1 THEN DO;  
    IF pillfreq2=0 THEN adhrate = (0.5 + 0.0) / 14;          /* not at all          */  
    IF pillfreq2=1 THEN adhrate = (0.5 + 0.5) / 14;          /* < 1 day per week   */  
    IF pillfreq2=2 THEN adhrate = (0.5 + 1.5) / 14;          /* 1-2 days per week  */  
    IF pillfreq2=3 THEN adhrate = (0.5 + 3.5) / 14;          /* 3-4 days per week  */  
    IF pillfreq2=4 THEN adhrate = (0.5 + 5.5) / 14;          /* 5-6 days per week  */  
    IF pillfreq2=5 THEN adhrate = (0.5 + 7.0) / 14;          /* every day of the week */  
    IF pillfreq2=. THEN adhrate = (0.5 + 7*d) / 14;          /* missing             */  
  END;  
  IF pillfreq1=2 THEN DO;  
    IF pillfreq2=0 THEN adhrate = (1.5 + 0.0) / 14;          /* not at all          */  
    IF pillfreq2=1 THEN adhrate = (1.5 + 0.5) / 14;          /* < 1 day per week   */  
    IF pillfreq2=2 THEN adhrate = (1.5 + 1.5) / 14;          /* 1-2 days per week  */  
    IF pillfreq2=3 THEN adhrate = (1.5 + 3.5) / 14;          /* 3-4 days per week  */  
    IF pillfreq2=4 THEN adhrate = (1.5 + 5.5) / 14;          /* 5-6 days per week  */  
    IF pillfreq2=5 THEN adhrate = (1.5 + 7.0) / 14;          /* every day of the week */  
    IF pillfreq2=. THEN adhrate = (1.5 + 7*d) / 14;          /* missing             */  
  END;  
  IF pillfreq1=3 THEN DO;  
    IF pillfreq2=0 THEN adhrate = (3.5 + 0.0) / 14;          /* not at all          */  
    IF pillfreq2=1 THEN adhrate = (3.5 + 0.5) / 14;          /* < 1 day per week   */  
    IF pillfreq2=2 THEN adhrate = (3.5 + 1.5) / 14;          /* 1-2 days per week  */  
    IF pillfreq2=3 THEN adhrate = (3.5 + 3.5) / 14;          /* 3-4 days per week  */  
    IF pillfreq2=4 THEN adhrate = (3.5 + 5.5) / 14;          /* 5-6 days per week  */  
    IF pillfreq2=5 THEN adhrate = (3.5 + 7.0) / 14;          /* every day of the week */  
    IF pillfreq2=. THEN adhrate = (3.5 + 7*d) / 14;          /* missing             */  
  END;  
  IF pillfreq1=4 THEN DO;  
    IF pillfreq2=0 THEN adhrate = (5.5 + 0.0) / 14;          /* not at all          */  
    IF pillfreq2=1 THEN adhrate = (5.5 + 0.5) / 14;          /* < 1 day per week   */  
    IF pillfreq2=2 THEN adhrate = (5.5 + 1.5) / 14;          /* 1-2 days per week  */  
    IF pillfreq2=3 THEN adhrate = (5.5 + 3.5) / 14;          /* 3-4 days per week  */  
    IF pillfreq2=4 THEN adhrate = (5.5 + 5.5) / 14;          /* 5-6 days per week  */  
    IF pillfreq2=5 THEN adhrate = (5.5 + 7.0) / 14;          /* every day of the week */  
    IF pillfreq2=. THEN adhrate = (5.5 + 7*d) / 14;          /* missing             */  
  END;  
  IF pillfreq1=5 THEN DO;  
    IF pillfreq2=0 THEN adhrate = (7.0 + 0.0) / 14;          /* not at all          */  
    IF pillfreq2=1 THEN adhrate = (7.0 + 0.5) / 14;          /* < 1 day per week   */  
    IF pillfreq2=2 THEN adhrate = (7.0 + 1.5) / 14;          /* 1-2 days per week  */  
    IF pillfreq2=3 THEN adhrate = (7.0 + 3.5) / 14;          /* 3-4 days per week  */  
    IF pillfreq2=4 THEN adhrate = (7.0 + 5.5) / 14;          /* 5-6 days per week  */  
    IF pillfreq2=5 THEN adhrate = (7.0 + 7.0) / 14;          /* every day of the week */  
    IF pillfreq2=. THEN adhrate = (7.0 + 7*d) / 14;          /* missing             */  
  END;  
  IF pillfreq1=. THEN DO;  
    IF pillfreq2=0 THEN adhrate = (7*d + 0.0) / 14;          /* not at all          */  
    IF pillfreq2=1 THEN adhrate = (7*d + 0.5) / 14;          /* < 1 day per week   */  
    IF pillfreq2=2 THEN adhrate = (7*d + 1.5) / 14;          /* 1-2 days per week  */  
    IF pillfreq2=3 THEN adhrate = (7*d + 3.5) / 14;          /* 3-4 days per week  */  
    IF pillfreq2=4 THEN adhrate = (7*d + 5.5) / 14;          /* 5-6 days per week  */  
    IF pillfreq2=5 THEN adhrate = (7*d + 7.0) / 14;          /* every day of the week */  
    IF pillfreq2=. THEN adhrate = d;  
  END;  
END;  
  
IF 0<adhrate<=0.01 THEN adhrate=0;          /* if taking <1% in a given year, then recode as taking none */  
IF adhrate> 1 THEN adhrate=1;              /* if taking >100% in a given year, then recode as taking one */
```

```

/*-----
  ASSIGN DOSE (INCLUDING NON-STUDY HORMONE USE
-----*/
extra=0; dose_e=0;
IF (estrogen=1 & progest=1) OR estpro=1 OR openlabel=1 THEN DO;
  extra=1; dose_e=RANUNI(2321);
END;

IF r=1 THEN DO;
  IF extra=0 THEN dose=adhrate;
  IF extra=1 THEN dose=adhrate + dose_e;
END;

IF r=0 THEN DO;
  IF extra=0 THEN dose=0;
  IF extra=1 THEN dose=dose_e;
END;

/*-----
  CREATE VARIABLES FOR WEIGHT ESTIMATION
-----*/
IF dose=0 THEN cendose=0; ELSE cendose=1; /* cendose: censoring indicator for those w/ dose=0 */
IF dose>0 THEN logdose=LOG(dose); /* logdose: log-transformed dose for those w/ dose>0 */
ldose = LAG(dose); IF FIRST.id THEN ldose=dose; /* hormone dose in the previous year */
output ipw0;

DROP adhrate_old d;
RUN;

%let _dsid_ = %sysfunc(open(_idtmp_));
%let _numids_ = %sysfunc(attrn(&_dsid_,nobs));
%let _rc_ = %sysfunc(close(&_dsid_));

%if &bootstrap = 1 %then %do;
  data _idholders_ (index = (sample));
  do sample = 0 to &nboot ;
    do newid = 1 to &_numids_;
      output ;
    end;
  end;
run;

proc surveysselect data= _idholders_
  method = urs
  n= &_numids_
  seed = 1232
  out = _idsamples (keep = sample newid numberhits rename = (numberhits = hits0))
  outall
;
strata sample ;
run;

data sample0 ;
set _idsamples ;
by sample ;
if sample = 0 then hits0 = 1; /* redefine numberhits in base sample */
retain newid2 ;
if first.sample then newid2 = 0;
do person = 1 to hits0 ;
  newid2=newid2+1 ;
output;
end;
run;

proc surveysselect data=sample0
  method=urs
  n = 100000
  seed = 98769
  out = sample01 (keep = sample newid newid2 hits0 numberhits rename= (numberhits = hits1))
;
strata sample ;
run;

proc sql ;
create table sample02 as
select sample,newid,newid2,hits0 , sum(hits1) as sum1
from sample01
group by sample, newid
;
quit;

proc sort data= sample02 nodupkey ;

```

```

    by sample newid ;
    run;

    data _idsamples;
    merge _idsamples(rename = (hits0 = numberhits )) /* original bootstrap */
          sample02 (rename = (sum1 = numberhits2)); /* for simulated data from bootstrap sample */
    by sample newid ;
    if numberhits2 = . then numberhits2=0;
    if sample = 0 then numberhits = 1 ; /* redefine numberhits in base sample */
    run;

%end;

%mend;

%macro getbootstrap_sample (datain = ,
                           dataout = ,
                           bsample = );

    data _idsneeded_ ;
    set _idsamples (where = (sample = %eval(&bsample))) ;
    run;

    data &dataout ;
    merge &datain _idsneeded_ (keep = newid numberhits ) ;
    by newid ;
    if numberhits = 0 then delete ;

    run;

%mend ;

%macro calculate_weights;

/*****
ESTIMATE WEIGHT
*****/
/*-----
PROBABILITY OF HAVING DOSE=0
-----*/
/* model 1a - numerator of sw */
PROC LOGISTIC DATA=ipw(WHERE=( r=0));
MODEL cendose=year year*year
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      ldose ldose*ldose ldose*ldose*ldose / ridging = none ;
freq numberhits;
OUTPUT OUT=modell1a P=num_adh;
RUN;

PROC LOGISTIC DATA=ipw(WHERE = (r = 1));
MODEL cendose=year year*year
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      ldose ldose*ldose ldose*ldose*ldose / ridging = none ;
freq numberhits;
OUTPUT OUT=modell2a P=num_adh;
RUN;

/* model 1b - denominator of sw */
PROC LOGISTIC DATA=ipw(WHERE = (r=0));
MODEL cendose=year year*year
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
      oc brcafrel bkbone1
      educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
      menarche1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
      region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
      physfunb2--physfunb4 preventb1--preventb2
      chdb cardiob cancerb bkboneb osteopb

      lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
      looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
      lbrstexam lmmamogrm lbrstbpsy lbrmafin0--lbrmafin2
      lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
      lchd lcardio1--lcardio3 lcancer lfracadj lsteop
      lvagbleed lheavybld1--lheavybld3 lintermit lbleednow

```

```

    lbrsttend lsevtend1--lsevtend2 lbrstchnng
    ldose ldose*ldose ldose*ldose*ldose / ridging = none ;
freq numberhits;
OUTPUT OUT=model1b1 P=den_adh;
RUN;

PROC LOGISTIC DATA=ipw(WHERE = (r=1));
MODEL cendose=year year*year
    dmflag ageg2--ageg6 ethnic3--ethnic8
    bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
    parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
    oc brcafrel bkbnonrel
    educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
    menarche1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
    region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
    physfunb2--physfunb4 preventb1--preventb2
    chdb cardiob cancerb bkboneb osteopb

    lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
    looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
    lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
    lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
    lchd lcardio1--lcardio3 lcancer lfracadj lsteop
    lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
    lbrsttend lsevtend1--lsevtend2 lbrstchnng
    ldose ldose*ldose ldose*ldose*ldose / ridging = none ;
freq numberhits;
OUTPUT OUT=model1b2 P=den_adh;
RUN;

/*-----
PROBABILITY OF HAVING A PARTICULAR DOSE AMONG THOSE W/ DOSE>0
-----*/
/* model 2a - numerator of sw */
PROC GLM DATA=ipw(WHERE = (cendose=1 AND r=0))
    ORDER=FREQ OUTSTAT=sse2a1(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
    dmflag ageg2--ageg6 ethnic3--ethnic8
    bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
    parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
    oc brcafrel
    tothcat1--tothcat3
    menarche1--menarche2
    marital1--marital3 menoage1--menoage9
    ldose ldose*ldose ldose*ldose*ldose / solution;
freq numberhits;
OUTPUT OUT=model2a1_1 P=pred;
RUN;
QUIT;

PROC GLM DATA=ipw(WHERE = ( cendose=1 AND r=1))
    ORDER=FREQ OUTSTAT=sse2a2(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
    dmflag ageg2--ageg6 ethnic3--ethnic8
    bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
    parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
    oc brcafrel
    tothcat1--tothcat3
    menarche1--menarche2
    marital1--marital3 menoage1--menoage9
    ldose ldose*ldose ldose*ldose*ldose / solution;
freq numberhits;
OUTPUT OUT=model2a2_1 P=pred;
RUN;
QUIT;

/* MODEL 2b - denominator of sw */
PROC GLM DATA=ipw(WHERE = ( cendose=1 AND r=0))
    ORDER=FREQ OUTSTAT=sse2b1(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
    dmflag ageg2--ageg6 ethnic3--ethnic8
    bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
    parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
    oc brcafrel bkbnonrel
    educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
    menarche1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
    region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
    physfunb2--physfunb4 preventb1--preventb2
    chdb cardiob cancerb bkboneb osteopb

    lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
    looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
    lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
    lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
    lchd lcardio1--lcardio3 lcancer lfracadj lsteop
    lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
    lbrsttend lsevtend1--lsevtend2 lbrstchnng
    ldose ldose*ldose ldose*ldose*ldose / solution;
freq numberhits;
OUTPUT OUT=model2b1_1 P=pred R=residual;

```

```

RUN;
QUIT;

PROC GLM DATA=ipw(WHERE = ( cendose=1 AND r=1))
  ORDER=FREQ OUTSTAT=sse2b2(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
  oc brcafrel bkbonrel
  educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
  menarche1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
  region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoaage1--menoaage9
  physfunb2--physfunb4 preventb1--preventb2
  chdb cardiob cancerb bkboneb osteopb

  lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
  looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
  lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
  lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
  lchd lcardio1--lcardio3 lcancer lfracadj losteop
  lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
  lbrsttend lsevtend1--lsevtend2 lbrstchnge
  ldose ldose*ldose ldose*ldose*ldose / solution;

freq numberhits;
OUTPUT OUT=model2b2_1 P=pred R=residual;
RUN;
QUIT;

DATA sse2a1;
SET sse2a1;
rootmse_n=SQRT(ss/df);
RUN;

DATA sse2a2;
SET sse2a2;
rootmse_n=SQRT(ss/df);
RUN;

DATA sse2b1;
SET sse2b1;
rootmse_n=SQRT(ss/df);
RUN;

DATA sse2b2;
SET sse2b2;
rootmse_n=SQRT(ss/df);
RUN;

PROC SQL;
  CREATE TABLE model2a1 AS
  SELECT O.*, L.rootmse_n
  FROM model2a1_1 O, sse2a1 L;
QUIT;

PROC SQL;
  CREATE TABLE model2a2 AS
  SELECT O.*, L.rootmse_n
  FROM model2a2_1 O, sse2a2 L;
QUIT;

PROC SQL;
  CREATE TABLE model2b1 AS
  SELECT O.*, L.rootmse_n
  FROM model2b1_1 O, sse2b1 L;
QUIT;

PROC SQL;
  CREATE TABLE model2b2 AS
  SELECT O.*, L.rootmse_n
  FROM model2b2_1 O, sse2b2 L;
QUIT;

DATA model2a1;
SET model2a1;
num_adhrate=PDF('NORMAL', logdose, pred, rootmse_n);
FORMAT num_adhrate 12.10;
RUN;

DATA model2a2;
SET model2a2;
num_adhrate=PDF('NORMAL', logdose, pred, rootmse_n);
FORMAT num_adhrate 12.10;
RUN;

DATA model2b1;
SET model2b1;
den_adhrate=PDF('NORMAL', logdose, pred, rootmse_n);
FORMAT den_adhrate 12.10;
RUN;

```

```
DATA model2b2;
SET model2b2;
den_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);
FORMAT den_adhrate 12.10;
RUN;
```

```
/*-----
ESTIMATE STABILIZED WEIGHT
-----*/
```

```
DATA ipw;
SET ipw (KEEP= numberhits newid id r adhrate dose dose_e extra cendose year invasive
invayear invady death dyear deathdy lastcont ager menoage dmflag age ageg2--ageg6
ethnic3--ethnic8 bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
parity1--parity9 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2 oc brcafrel tothcat
tothcat1--tothcat3 menarche1--menarche2 marital1--marital3 menoage1--menoage9);
RUN;
```

```
DATA modell1a1;
SET modell1a1 (KEEP=id year num_adh);
RUN;
```

```
DATA modella2;
SET modella2 (KEEP=id year num_adh);
RUN;
```

```
DATA modell1b1;
SET modell1b1 (KEEP=id year den_adh);
RUN;
```

```
DATA modell1b2;
SET modell1b2 (KEEP=id year den_adh);
RUN;
```

```
DATA model2a1;
SET model2a1 (KEEP=id year num_adhrate);
RUN;
```

```
DATA model2a2;
SET model2a2 (KEEP=id year num_adhrate);
RUN;
```

```
DATA model2b1;
SET model2b1 (KEEP=id year den_adhrate);
RUN;
```

```
DATA model2b2;
SET model2b2 (KEEP=id year den_adhrate);
RUN;
```

```
PROC SORT TAGSORT DATA=ipw;
BY id year;
RUN;
```

```
PROC SORT TAGSORT DATA=modell1a1;
BY id year;
RUN;
```

```
PROC SORT TAGSORT DATA=modella2;
BY id year;
RUN;
```

```
PROC SORT TAGSORT DATA=modell1b1;
BY id year;
RUN;
```

```
PROC SORT TAGSORT DATA=modell1b2;
BY id year;
RUN;
```

```
PROC SORT TAGSORT DATA=model2a1;
BY id year;
RUN;
```

```
PROC SORT TAGSORT DATA=model2a2;
BY id year;
RUN;
```

```
PROC SORT TAGSORT DATA=model2b1;
BY id year;
RUN;
```

```
PROC SORT TAGSORT DATA=model2b2;
BY id year;
RUN;
```

```
DATA main;
MERGE ipw modell1a1 modella2 modell1b1 modell1b2 model2a1 model2a2 model2b1 model2b2;
BY id year;
IF FIRST.id THEN stabw=1;
RETAIN stabw;
```



```
w2 = num_adh / den_adh;
w3 = (1 - num_adh) / (1 - den_adh);
w4 = num_adhrate / den_adhrate;
```

```
IF cendose=0 THEN stabw = stabw * w2;
IF cendose=1 THEN stabw = stabw * w3 * w4;
```

```
DROP w2-w4; RUN;
```

```
PROC SQL;
CREATE TABLE a AS
SELECT id, MAX(year) AS maxyear
FROM main GROUP BY id;
QUIT;
```

```
PROC SQL;
CREATE TABLE b AS
SELECT O.*, L.maxyear
FROM main O, a L
WHERE O.id=L.id;
QUIT;
```

```
DATA c;
SET b;
invamo = CEIL(invady/30.4375);
deathmo = CEIL(deathdy/30.4375);
```

```
mm=0;
RETAIN mm;
DO y=1 UNTIL (year=maxyear);
DO m=1 TO 12;
mm=mm+1;
OUTPUT;
END;
END;
DROP y m;
RUN;
```

```
DATA d; SET c;
month=mm+12*(year-1);
```

```
maxmo = maxyear*12;
lmo = CEIL(lastcont/30.4375);
lyear = CEIL(lastcont/365.25);
lastmo = MIN(lmo,maxmo); /* lmo: last month based on lastcont; maxmo: last month based on data */
lastyear = MIN(lyear,maxyear); /* lyear: last year based on lastcont; maxyear: last year based on data */
```

```
IF invamo NE . AND month > invamo THEN DELETE;
IF deathmo NE . AND month > deathmo THEN DELETE;
IF invamo = . AND deathmo = . AND month > lastmo THEN DELETE;
```

```
IF invamo NE . AND invamo = month THEN invasive=1;
ELSE invasive=0;
IF deathmo NE . AND deathmo = month THEN death=1;
ELSE death=0;
```

```
DROP mm;
RUN;
```

```
/*-----
CALCULATE CUMULATIVE USE
-----*/
```

```
DATA p;
SET whi.baseline_bc(KEEP=id);
p=RANUNI(99999); RUN;
```

```
DATA e;
MERGE p d (in = a);
BY id;
if a ;
RUN;
```

```
PROC SORT TAGSORT DATA=e;
BY id month;
RUN;
```

```
DATA dr_bc(KEEP= numberhits cuma stabw id newid r month year invasive ager menoage dmflag
ageg2--ageg6 ethnic3--ethnic8 bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2 oc brcafel tothcat tothcat1--tothcat3
menarche1--menarche2 marital1--marital3 menoage1--menoage9);
```

```
SET e;
BY id month;
```

```
IF FIRST.id THEN cuma = 0;
```

```

RETAIN cuma;

IF (year NE lastyear) & (year NE invayear) & (year NE dyear) THEN DO;
  cuma = cuma + (dose / 12);
END;

IF (year=lastyear) & (year NE invayear) & (year NE dyear) THEN DO;
  d0 = (lastmo-((lastyear-1)*12));
  cuma = cuma + (dose / d0);
END;

IF year = dyear THEN DO;
  d1 = (deathmo-((dyear-1)*12));
  cuma = cuma + (dose / d1);
END;

IF (year = invayear) & (invayear NE lastyear) & (year NE dyear) THEN DO;
  d2 = invamo-((invayear-1)*12);
  d3 = 13 - d2;
  IF 0/d3 <= p <= 1/d3 THEN cuma = cuma + (dose / d2);
  IF 1/d3 < p <= 2/d3 THEN cuma = cuma + (dose / (d2+1));
  IF 2/d3 < p <= 3/d3 THEN cuma = cuma + (dose / (d2+2));
  IF 3/d3 < p <= 4/d3 THEN cuma = cuma + (dose / (d2+3));
  IF 4/d3 < p <= 5/d3 THEN cuma = cuma + (dose / (d2+4));
  IF 5/d3 < p <= 6/d3 THEN cuma = cuma + (dose / (d2+5));
  IF 6/d3 < p <= 7/d3 THEN cuma = cuma + (dose / (d2+6));
  IF 7/d3 < p <= 8/d3 THEN cuma = cuma + (dose / (d2+7));
  IF 8/d3 < p <= 9/d3 THEN cuma = cuma + (dose / (d2+8));
  IF 9/d3 < p <= 10/d3 THEN cuma = cuma + (dose / (d2+9));
  IF 10/d3 < p <= 11/d3 THEN cuma = cuma + (dose / (d2+10));
  IF 11/d3 < p <= 12/d3 THEN cuma = cuma + (dose / (d2+11));
END;

IF (year = invayear) & (invayear=lastyear) & (year NE dyear) THEN DO;
  d4 = invamo-((invayear-1)*12);
  d5 = lastmo-((lastyear-1)*12);
  d6 = (d5-d4) + 1;

  IF 0/d6 <= p <= 1/d6 THEN cuma = cuma + (dose / d4);
  IF 1/d6 < p <= 2/d6 THEN cuma = cuma + (dose / (d4+1));
  IF 2/d6 < p <= 3/d6 THEN cuma = cuma + (dose / (d4+2));
  IF 3/d6 < p <= 4/d6 THEN cuma = cuma + (dose / (d4+3));
  IF 4/d6 < p <= 5/d6 THEN cuma = cuma + (dose / (d4+4));
  IF 5/d6 < p <= 6/d6 THEN cuma = cuma + (dose / (d4+5));
  IF 6/d6 < p <= 7/d6 THEN cuma = cuma + (dose / (d4+6));
  IF 7/d6 < p <= 8/d6 THEN cuma = cuma + (dose / (d4+7));
  IF 8/d6 < p <= 9/d6 THEN cuma = cuma + (dose / (d4+8));
  IF 9/d6 < p <= 10/d6 THEN cuma = cuma + (dose / (d4+9));
  IF 10/d6 < p <= 11/d6 THEN cuma = cuma + (dose / (d4+10));
  IF 11/d6 < p <= 12/d6 THEN cuma = cuma + (dose / (d4+11));
END;
RUN;

%mend;

%macro numargs(arg);

  %let n = 1;
  %if &arg^= %then %do;
    %do %until (%scan(&arg,%eval(&n),%str( ))=%str());

      %let word = %scan(&arg,&n);
      %let n = %eval(&n+1);

    %end;

  %end;
  %eval(&n-1) /* there is no ; here since it will be used as %let a = %numargs(&b) ;
              and the ; is included at the end of this line */
%mend numargs;

%MACRO table4 ;

data seeds0;
set whi.seeds(where = (bsample = %eval(&bsample+1)));
%do _i_ = 1 %to 7;
  call symput("seed&i_",trim(left(seed&i_)));
%end;
run;

/* at this point in the program dr_bc only contains subjects who
   have numberhits > 0 */

data tmp ;
set _idsneeded_(where = (numberhits > 0)) ;

```

```
run;
```

```
data dr_bc;  
merge dr_bc tmp (keep = newid numberhits2) ;  
by newid ;
```

```
run;
```

```
/* create a new data set that has one observation for each subject that will  
contribute to the simulated data set used for the hazard ratio model. This data  
is constructed as follows:  
1) a random sample using numberhits1 > 0 was selected for the bootstrap sample  
from the original 16,000 subjects.  
2) for each subject with numberhits > 1 we expanded this subject into numberhits new  
subjects, and then a new random sample of 100,000 subjects was selected. The number of  
time each of these new subjects was selected was summed over those that came from the  
same original subject. This is numberhits2 which is the number of new subjects that  
will be included in the simulated data set.  
*/
```

```
*/
```

```
data test(compress = yes) ;  
set dr_bc (where = ( month = 1) keep = numberhits2  
month ager tothcat menoage  
dmflag ageg2--ageg6 ethnic3--ethnic8  
bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9  
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2  
oc brcafrel  
tothcat1--tothcat3  
menarche1--menarche2  
marital1--marital3 menoage1--menoage9 );
```

```
retain newid ;  
if _n_ = 1 then newid = 0 ;
```

```
do person = 1 to numberhits2;  
newid = newid + 1 ;  
output ;  
end;  
run;
```

```
/** models for ageg = 1 **/
```

```
DATA r;  
SET DR_BC;  
%RCSPLINE(month,6,48,90);  
use1 = cuma * (ager=1);  
use2 = cuma * (ager in (2,3));
```

```
interact1 = use1 * month;  
interact2 = use1 * month1;  
interact3 = use2 * month;  
interact4 = use2 * month1;
```

```
a = (tothcat ne 0) * month;  
b = (tothcat ne 0) * month1;  
RUN;
```

```
/* PROC LOGISTIC generated the same curve as PROC GENMOD */  
PROC LOGISTIC DESCENDING DATA=R;  
ODS OUTPUT ParameterEstimates=PE;  
MODEL Invasive = use1-use2  
month month1  
dmflag ageg2--ageg6 ethnic3--ethnic8  
bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9  
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2  
oc brcafrel  
tothcat1--tothcat3  
menarche1--menarche2  
marital1--marital3 menoage1--menoage9  
interact1-interact4 a b ;  
FREQ numberhits ;  
WEIGHT stabw;  
RUN;
```

```
/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */  
PROC SQL NOPRINT;  
SELECT estimate FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' '  
FROM pe;  
QUIT;
```

```
PROC SQL NOPRINT;  
SELECT variable INTO:model SEPARATED BY ' '  
FROM pe;  
QUIT;
```

```

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=(_STAT_="N"));
RUN;

PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' '
FROM nobs;
QUIT;

DATA r1 ;
SET test( where = ( ager=1 ));
length event treated 3 month 5;
ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

retain seed ;
if _n_ = 1 then seed = &seed1 ;* 123456;

intercept = 1;

event0 = 0;
event1 = 0 ;

untreated_done = 0;
treated_done = 0 ;

DO month=1 TO 96;
  %RCSPLINE(month,6,48,90);
  xbeta0 = 0;
  xbetal = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  do i=1 to DIM(var);
    use1      = 0;
    use2      = 0;
    use3      = 0;
    use4      = 0;
    interact1 = 0;
    interact2 = 0;
    interact3 = 0;
    interact4 = 0;

    xbeta0    = xbeta0 + coef[i] * var[i];

    use1      = (month/12) * (ager=1);
    use2      = (month/12) * (ager IN (2,3));

    interact1 = use1 * month;
    interact2 = use1 * month1;

    interact3 = use2 * month;
    interact4 = use2 * month1;

    xbetal    = xbetal + coef[i] * var[i];
  end;

  p0 = 1 / (1 + EXP(-xbeta0));
  p1 = 1 / (1 + EXP(-xbetal));

  if untreated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p0,event0);
    if event0 = 1 then untreated_done = 1;
    event = event0;
    treated = 0 ;
    output ;
  end;
  if treated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p1,event1);
    if event1 = 1 then treated_done = 1 ;
    event = event1;
    treated = 1 ;
    output ;
  end;

END;

drop event0 event1 p0 p1 xbeta0 xbetal seed intercept ;
RUN;

```

```

PROC LOGISTIC DESCENDING DATA=R1 (where = (month <= 24));
ODS OUTPUT ParameterEstimates=PE24;
MODEL event = treated
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafre1
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      a b ;

RUN;

PROC LOGISTIC DESCENDING DATA=R1 (where = (month <= 72));
ODS OUTPUT ParameterEstimates=PE72;
MODEL event = treated
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafre1
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      a b ;

*freq hits1;
RUN;

PROC LOGISTIC DESCENDING DATA=R1;
ODS OUTPUT ParameterEstimates=PE96;
MODEL event = treated
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafre1
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      a b ;

*freq hits1 ;
RUN;

data MEAN_FOR_model6 ;
merge pe24 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta24))
      pe72 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta72))
      pe96 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta96)) ;
bsample = &bsample ;
drop variable ;
run;

proc datasets library = work nolist ;
delete r r1 pe24 pe72 pe96;
quit;

/** model for ageg = 2 or ageg = 3 ***/

DATA r;
SET dr_bc;
%RCSPLINE(month,6,48,90);
use1      = cuma      * (ager=1);
use2      = cuma      * (ager IN (2,3));

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
RUN;

PROC LOGISTIC DESCENDING DATA=r ;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use2
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafre1

```

```

        tothcat1--tothcat3
        menarchel--menarche2
        marital1--marital3 menoage1--menoage9
        interact1-interact4 a b ;

freq numberhits ;
WEIGHT stabw;
RUN;

/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;

PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=(_STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nob;
QUIT;

DATA r1 ;
SET test( WHERE = ( AGER IN (2,3)));
length event treated 3 month 5;

ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

intercept = 1;
if _n_ = 1 then seed = &seed2 ; *123456;

untreated_done = 0;
treated_done = 0 ;

event0 = 0;
event1 = 0 ;

retain seed ;

DO month=1 TO 96;
    %RCSPLINE(month,6,48,90);
    xbeta0 = 0;
    xbeta1 = 0;
    a = (tothcat NE 0) * month;
    b = (tothcat NE 0) * month1;

DO i=1 TO DIM(var);
    use1 = 0; /* never treated */
    use2 = 0;

    interact1 = 0;
    interact2 = 0;
    interact3 = 0;
    interact4 = 0;

    xbeta0 = xbeta0 + coef[i] * var[i];

    use1 = (month/12) * (ager=1); /* always treated */
    use2 = (month/12) * (ager IN (2,3));

    interact1 = use1 * month;
    interact2 = use1 * month1;

    interact3 = use2 * month;
    interact4 = use2 * month1;

    xbeta1 = xbeta1 + coef[i] * var[i];
end;

p0 = 1 / (1 + EXP(-xbeta0));
p1 = 1 / (1 + EXP(-xbeta1));

if untreated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p0,event0);

```

```

        if event0 = 1 then untreated_done = 1;
        event = event0;
        treated = 0 ;
        output ;
    end;
    if treated_done = 0 then do ;
        event = . ;
        call ranbin(seed,1,p1,event1);
        if event1 = 1 then treated_done = 1;
        event = event1;
        treated = 1 ;
        output ;
    end;
END;

drop event0 event1 p0 p1 xbeta0 xbeta1 seed intercept ;
RUN;

```

```

PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 24));
ODS OUTPUT ParameterEstimates=PE24;
MODEL event = treated
    month month1
        dmflag ageg2--ageg6 ethnic3--ethnic8
        bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
        parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
        oc brcafrel
        tothcat1--tothcat3
        menarche1--menarche2
        marital1--marital3 menoage1--menoage9
        a b ;
run;

```

```

PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 72));
ODS OUTPUT ParameterEstimates=PE72;
MODEL event = treated
    month month1
        dmflag ageg2--ageg6 ethnic3--ethnic8
        bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
        parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
        oc brcafrel
        tothcat1--tothcat3
        menarche1--menarche2
        marital1--marital3 menoage1--menoage9
        a b ;
run;

```

```

PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 96));
ODS OUTPUT ParameterEstimates=PE96;
MODEL event = treated
    month month1
        dmflag ageg2--ageg6 ethnic3--ethnic8
        bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
        parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
        oc brcafrel
        tothcat1--tothcat3
        menarche1--menarche2
        marital1--marital3 menoage1--menoage9
        a b ;
run;

```

```

data MEAN_FOR_model7 ;
merge pe24 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta24))
      pe72 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta72))
      pe96 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta96)) ;
bsample = &bsample ;
drop variable ;
run;

```

```

proc datasets library = work nolist ;
delete r r1 pe24 pe72 pe96;
quit;

```

```

/**** model for menoage = 0 ****/

```

```

DATA r;
SET dr_bc;
%RCSPLINE(month,6,48,90);
use1      = cuma      * (menoage=0);

```

```

use2      = cuma          * (menoage IN (1,2));
use3      = cuma          * (menoage=9);

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

interact5 = use3 * month;
interact6 = use3 * month1;

a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;

run;

PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use3
      month month1
      dmflag age2--age6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      interact1-interact6 a b ;

freq numberhits ;
WEIGHT stabw;
RUN;

/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;

PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM PE;
QUIT;

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=NOBS(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nob;
QUIT;

DATA R1 ;
SET test (WHERE = ( menoage=0));
length event treated 3 month 5;
ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

retain seed;
if _n_ = 1 then seed = &seed3 ; * 123456;

intercept = 1;

untreated_done = 0 ;
treated_done = 0 ;
event0 = 0;
event1 = 0 ;

DO month=1 TO 96;
  %RCSPLINE(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

DO i=1 TO DIM(var);
  use1      = 0;
  use2      = 0;
  use3      = 0;
  interact1 = 0;
  interact2 = 0;
  /* never treated */

```



```
interact3 = 0;
interact4 = 0;
interact5 = 0;
interact6 = 0;
```

```
xbeta0 = xbeta0 + coef[i] * var[i];
```

```
use1 = (month/12) * (menoage=0); /* always treated */
use2 = (month/12) * (menoage IN (1,2));
use3 = (month/12) * (menoage=9);
```

```
interact1 = use1 * month;
interact2 = use1 * month1;
```

```
interact3 = use2 * month;
interact4 = use2 * month1;
```

```
interact5 = use3 * month;
interact6 = use3 * month1;
```

```
xbetal = xbetal + coef[i] * var[i];
end;
```

```
p0 = 1 / (1 + EXP(-xbeta0));
p1 = 1 / (1 + EXP(-xbetal));
```

```
if untreated_done = 0 then do;
  event = . ;
  call ranbin(seed,1,p0,event0);
  if event0 = 1 then untreated_done = 1 ;
  event = event0;
  treated = 0 ;
  output ;
end;
```

```
if treated_done = 0 then do ;
  event = .;
  call ranbin(seed,1,p1,event1);
  if event1 = 1 then treated_done = 1;
  event = event1;
  treated = 1 ;
  output ;
end;
```

END;

```
drop event0 event1 p0 p1 xbeta0 xbetal seed intercept ;
RUN;
```

```
PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 24));
ODS OUTPUT ParameterEstimates=PE24;
MODEL event = treated
  month month1
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafre1
  tothcat1--tothcat3
  menarche1--menarche2
  marital1--marital3 menoage1--menoage9
  a b ;
```

run;

```
PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 72));
ODS OUTPUT ParameterEstimates=PE72;
MODEL event = treated
  month month1
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafre1
  tothcat1--tothcat3
  menarche1--menarche2
  marital1--marital3 menoage1--menoage9
  a b ;
```

run;

```
PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 96));
ODS OUTPUT ParameterEstimates=PE96;
MODEL event = treated
  month month1
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
```

```
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
oc brcafrel
tothcat1--tothcat3
menarche1--menarche2
marital1--marital3 menoage1--menoage9
a b ;
```

```
run;
```

```
data MEAN_FOR_model4 ;
merge pe24 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta24))
      pe72 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta72))
      pe96 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta96)) ;
bsample = &bsample ;
drop variable ;
run;
```

```
proc datasets library = work nolist ;
delete r r1 pe24 pe72 pe96;
quit;
```

```
/**** model for menoage = 1 or menoage = 2 *****/
```

```
DATA r;
SET dr_bc;
%RCSPLINE(month,6,48,90);
use1      = cuma      * (menoage=0);
use2      = cuma      * (menoage IN (1,2));
use3      = cuma      * (menoage=9);
```

```
interact1 = use1 * month;
interact2 = use1 * month1;
```

```
interact3 = use2 * month;
interact4 = use2 * month1;
```

```
interact5 = use3 * month;
interact6 = use3 * month1;
```

```
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
RUN;
```

```
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use3
              month month1
              dmflag ageg2--ageg6 ethnic3--ethnic8
              bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
              parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
              oc brcafrel
              tothcat1--tothcat3
              menarche1--menarche2
              marital1--marital3 menoage1--menoage9
              interact1-interact6 a b ;
```

```
freq numberhits ;
WEIGHT stabw;
RUN;
```

```
/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;
```

```
PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;
```

```
/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nobs;
QUIT;
```

```

DATA r1 ;
SET test (WHERE = ( menoage IN (1,2)));

length event treated 3 month 5;

ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

retain seed;
if _n_ = 1 then seed = &seed4 ; * 123456;

intercept = 1;

untreated_done = 0;
treated_done = 0 ;

event0 = 0;
event1 = 0 ;

DO month=1 TO 96;
  %RCSPLINE(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  DO i=1 TO DIM(var);
    use1 = 0; /* never treated */
    use2 = 0;
    use3 = 0;

    interact1 = 0;
    interact2 = 0;
    interact3 = 0;
    interact4 = 0;
    interact5 = 0;
    interact6 = 0;

    xbeta0 = xbeta0 + coef[i] * var[i];

    use1 = (month/12) * (menoage=0); /* always treated */
    use2 = (month/12) * (menoage IN (1,2));
    use3 = (month/12) * (menoage=9);

    interact1 = use1 * month;
    interact2 = use1 * month1;

    interact3 = use2 * month;
    interact4 = use2 * month1;

    interact5 = use3 * month;
    interact6 = use3 * month1;

    xbeta1 = xbeta1 + coef[i] * var[i];
  END;

  p0 = 1 / (1 + EXP(-xbeta0));
  p1 = 1 / (1 + EXP(-xbeta1));

  if untreated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p0,event0);
    if event0 = 1 then untreated_done = 1;
    event = event0;
    treated = 0 ;
    output ;
  end;
  if treated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p1,event1);
    if event1 = 1 then treated_done = 1 ;
    event = event1;
    treated = 1 ;
    output ;
  end;

end;
drop event0 event1 p0 p1 xbeta0 xbeta1 seed intercept ;
RUN;

```

```
PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 24));
ODS OUTPUT ParameterEstimates=PE24;
MODEL event = treated
```

```
    month month1
    dmflag ageg2--ageg6 ethnic3--ethnic8
    bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
    parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
    oc brcafrel
    tothcat1--tothcat3
    menarche1--menarche2
    marital1--marital3 menoage1--menoage9
    a b ;
```

```
run;
```

```
PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 72));
ODS OUTPUT ParameterEstimates=PE72;
MODEL event = treated
```

```
    month month1
    dmflag ageg2--ageg6 ethnic3--ethnic8
    bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
    parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
    oc brcafrel
    tothcat1--tothcat3
    menarche1--menarche2
    marital1--marital3 menoage1--menoage9
    a b ;
```

```
run;
```

```
PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 96));
ODS OUTPUT ParameterEstimates=PE96;
MODEL event = treated
```

```
    month month1
    dmflag ageg2--ageg6 ethnic3--ethnic8
    bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
    parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
    oc brcafrel
    tothcat1--tothcat3
    menarche1--menarche2
    marital1--marital3 menoage1--menoage9
    a b ;
```

```
run;
```

```
data MEAN_FOR_model5 ;
merge pe24 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta24))
      pe72 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta72))
      pe96 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta96)) ;
bsample = &bsample ;
drop variable ;
run;
```

```
proc datasets library = work nolist ;
delete r r1 pe24 pe72 pe96;
quit;
```

```
/** model for tothcat = 0 ***/
```

```
DATA r;
SET dr_bc;
%RCSPLINE(month,6,48,90);
use1      = cuma      * (tothcat=0)      ;
use2      = cuma      * (tothcat NE 0);
```

```
interact1 = use1 * month;
interact2 = use1 * month1;
```

```
interact3 = use2 * month;
interact4 = use2 * month1;
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
```

```
RUN;
```

```
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use2
```

```
    month month1
    dmflag ageg2--ageg6 ethnic3--ethnic8
    bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
    parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
```

```

        oc brcafrel
        tothcat1--tothcat3
        menarche1--menarche2
        marital1--marital3 menoage1--menoage9
        interact1-interact4 a b ;
FREQ numberhits ;
WEIGHT stabw;
RUN;

/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;

PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR DF;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nobs;
QUIT;

DATA r1 ;
SET test (WHERE= ( tothcat=0 ));

length event treated 3 month 5;
ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

retain seed ;
if _n_ = 1 then seed = &seed5 ; * 123456;

intercept = 1;

untreated_done = 0 ;
treated_done = 0 ;

event0 = 0;
event1 = 0 ;

DO month=1 TO 96;
  %RCSPLINE(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  DO i=1 TO DIM(var);
    use1 = 0; /* never treated */
    use2 = 0;

    interact1 = 0;
    interact2 = 0;
    interact3 = 0;
    interact4 = 0;

    xbeta0 = xbeta0 + coef[i] * var[i];

    use1 = (month/12) * (tothcat=0) ; /* always treated */
    use2 = (month/12) * (tothcat NE 0);

    interact1 = use1 * month;
    interact2 = use1 * month1;

    interact3 = use2 * month;
    interact4 = use2 * month1;

    xbeta1 = xbeta1 + coef[i] * var[i];
  END;

  p0 = 1 / (1 + EXP(-xbeta0));
  p1 = 1 / (1 + EXP(-xbeta1));

  if untreated_done = 0 then do ;
    event = . ;

```

```

    call ranbin(seed,1,p0,event0);
    if event0 = 1 then untreated_done = 1 ;
    event = event0;
    treated = 0 ;
    output ;
end;
if treated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p1,event1);
    if event1 = 1 then treated_done = 1 ;
    event = event1;
    treated = 1 ;
    output ;
end;

end;
drop event0 event1 p0 p1 xbeta0 xbeta1 seed intercept ;
RUN;

data fortest ;
set r1 (keep = event treated month where = (month<=24));
run;

proc sort data = fortest ;
by treated ;
run;

proc freq data = fortest ;
table month*event ;
by treated ;
run;

PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 24));
ODS OUTPUT ParameterEstimates=PE24;
MODEL event      = treated
            month month1
            dmflag ageg2--ageg6 ethnic3--ethnic8
            bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
            parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
            oc brcafre1
            tothcat1--tothcat3
            menarche1--menarche2
            marital1--marital3 menoage1--menoage9
            a b ;

run;

PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 72));
ODS OUTPUT ParameterEstimates=PE72;
MODEL event      = treated
            month month1
            dmflag ageg2--ageg6 ethnic3--ethnic8
            bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
            parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
            oc brcafre1
            tothcat1--tothcat3
            menarche1--menarche2
            marital1--marital3 menoage1--menoage9
            a b ;

run;

PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 96));
ODS OUTPUT ParameterEstimates=PE96;
MODEL event      = treated
            month month1
            dmflag ageg2--ageg6 ethnic3--ethnic8
            bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
            parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
            oc brcafre1
            tothcat1--tothcat3
            menarche1--menarche2
            marital1--marital3 menoage1--menoage9
            a b ;

run;

data MEAN_FOR_model2 ;
merge pe24 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta24))
      pe72 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta72))
      pe96 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta96)) ;
bsample = &bsample ;

```

```
drop variable ;
run;

PROC PRINT DATA = MEAN_FOR_MODEL7 ;
RUN;

proc datasets library = work nolist ;
delete r r1 pe24 pe72 pe96;
quit;
```

```
/** model for tothcat ^= 0 ***/
```

```
DATA r;
SET dr_bc;
%RCSPLINE(month,6,48,90);
use1      = cuma      * (tothcat=0) ;
use2      = cuma      * (tothcat NE 0);

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;
```

```
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
```

```
RUN;
```

```
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use2
              month month1
              dmflag ageg2--ageg6 ethnic3--ethnic8
              bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
              parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
              oc brcafre1
              tothcat1--tothcat3
              menarche1--menarche2
              marital1--marital3 menoage1--menoage9
              interact1-interact4 a b ;
```

```
freq numberhits ;
WEIGHT stabw;
RUN;
```

```
/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
```

```
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;
```

```
PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;
```

```
/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
```

```
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nob;
QUIT;
```

```
DATA r1 ;
SET test (WHERE= ( tothcat NE 0));
length event treated 3 month 5;
```

```
ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);
```

```
retain seed ;
if _n_ = 1 then seed = &seed6 ;*123456;
```

```
intercept = 1;
s0         = 1;
s1         = 1;
```

```

untreated_done = 0;
treated_done = 0 ;

event0 = 0;
event1 = 0 ;

DO month=1 TO 96;
  %RCSPLINE(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  DO i=1 TO DIM(var);
    use1      = 0;
    use2      = 0;
    interact1 = 0;
    interact2 = 0;
    interact3 = 0;
    interact4 = 0;

    xbeta0    = xbeta0 + coef[i] * var[i];

    use1      = (month/12)          * (tothcat=0) ;
    use2      = (month/12)          * (tothcat NE 0);
    interact1 = use1 * month;
    interact2 = use1 * month1;

    interact3 = use2 * month;
    interact4 = use2 * month1;

    xbeta1    = xbeta1 + coef[i] * var[i];
  END;

  p0 = 1 / (1 + EXP(-xbeta0));
  p1 = 1 / (1 + EXP(-xbeta1));

  if untreated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p0,event0);
    if event0 = 1 then untreated_done = 1;
    event = event0;
    treated = 0 ;
    output ;
  end;
  if treated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p1,event1);
    if event1 = 1 then treated_done = 1 ;
    event = event1;
    treated = 1 ;
    output ;
  end;

end;
drop event0 event1 p0 p1 xbeta0 xbeta1 seed intercept ;
RUN;

PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 24));
ODS OUTPUT ParameterEstimates=PE24;
MODEL event = treated
            month month1
            dmflag age2--age6 ethnic3--ethnic8
            bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
            parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
            oc brcafre1
            tothcat1--tothcat3
            menarche1--menarche2
            marital1--marital3 menoage1--menoage9
            a b ;

run;

PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 72));
ODS OUTPUT ParameterEstimates=PE72;
MODEL event = treated
            month month1
            dmflag age2--age6 ethnic3--ethnic8
            bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
            parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
            oc brcafre1
            tothcat1--tothcat3
            menarche1--menarche2
            marital1--marital3 menoage1--menoage9

```



a b ;

run;

```
PROC LOGISTIC DESCENDING DATA=r1 (where = (month <= 96));
ODS OUTPUT ParameterEstimates=PE96;
MODEL event = treated
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      a b ;
```

run;

```
data MEAN_FOR_model3 ;
merge pe24 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta24))
      pe72 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta72))
      pe96 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta96)) ;
bsample = &bsample ;
drop variable ;
run;
```

```
proc datasets library = work nolist ;
delete r r1 pe24 pe72 pe96;
quit;
```

/\* model for all women (invasive ^= .) \*/

```
%nopri :
DATA r;
SET dr_bc;
%RCSPLINE(month,6,48,90);
cuma2 = cuma * cuma;
interact1 = cuma * month;
interact2 = cuma * month1;
```

```
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
RUN;
```

```
/* PROC LOGISTIC generated the same curve as PROC GENMOD */
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=pe;
MODEL invasive = cuma
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      interact1-interact2 a b ;
FREQ numberhits ;
WEIGHT stabw;
RUN;
```

```
/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;
```

```
PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;
```

```
/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nobs;
QUIT;
```

```

DATA r1 ;
SET test ;
length event treated 3 month 5;

ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

retain seed ;
if _n_ = 1 then seed = &seed7 ; *123456;

intercept = 1;

untreated_done = 0 ;
treated_done = 0 ;
event0 = 0;
event1 = 0 ;

DO month=1 TO 96;
  %RCSPLINE(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  DO i=1 TO DIM(var);
    cuma = 0; /* never treated */

    interact1 = 0;
    interact2 = 0;

    xbeta0 = xbeta0 + coef[i] * var[i];

    cuma = (month/12); /* always treated */

    interact1 = (month/12) * month;
    interact2 = (month/12) * month1;

    xbeta1 = xbeta1 + coef[i] * var[i];
  END;

  p0 = 1 / (1 + EXP(-xbeta0));
  p1 = 1 / (1 + EXP(-xbeta1));

  if untreated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p0,event0);
    if event0 = 1 then untreated_done = 1;
    event = event0;
    treated = 0 ;
    output ;
  end;
  if treated_done = 0 then do ;
    event = . ;
    call ranbin(seed,1,p1,event1);
    if event1 = 1 then treated_done = 1 ;
    event = event1;
    treated = 1 ;
    output ;
  end;

end;

drop event0 event1 person_count0 person_count1 p0 p1 xbeta0 xbeta1 seed intercept ;
RUN;

```

```

PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 24));
ODS OUTPUT ParameterEstimates=pe24;
MODEL event = treated
  month month1
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafrel

```

```
tothcat1--tothcat3
menarche1--menarche2
marital1--marital3 menoage1--menoage9
a b ;
```

run;

```
PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 72));
ODS OUTPUT ParameterEstimates=pe72;
```

```
MODEL event = treated
  month month1
  dmflag age2--age6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafrel
  tothcat1--tothcat3
  menarche1--menarche2
  marital1--marital3 menoage1--menoage9
  a b ;
```

run;

```
PROC LOGISTIC DESCENDING DATA=r1(where = (month <= 96));
ODS OUTPUT ParameterEstimates=pe96;
```

```
MODEL event = treated
  month month1
  dmflag age2--age6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafrel
  tothcat1--tothcat3
  menarche1--menarche2
  marital1--marital3 menoage1--menoage9
  a b ;
```

run;

```
data MEAN_FOR_model1 ;
merge pe24 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta24))
      pe72 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta72))
      pe96 (keep = variable estimate where = ( variable = "treated") rename = (estimate = beta96)) ;
bsample = &bsample ;
drop variable ;
run;
```

```
proc datasets library = work nolist ;
delete r r1 pe24 pe72 pe96;
quit;
```

```
%MEND table4 ;
```

```
%macro results ;
```

```
%let cond1 = invasive ^= . ;
%let cond4 = menoage = 0 ;
%let cond5 = menoage =1 or menoage = 2 ;
%let cond6 = ager = 1 ;
%let cond7 = ager = 2 or ager = 3 ;
%let cond2 = tothcat = 0 ;
%let cond3 = tothcat ^= 0 ;
```

```
%do i = 1 %to 7;
```

```
data model&i ;
set whi.&&dataname.&i._0_200 (where = (beta24 ne .)) ;
run ;
```

```
proc transpose data = model&i(where = (bsample = 0))
  out = tcond&i (where = (_NAME_ ^= "bsample") rename = (COL1=Estimate)) ;
```

run;

```
title "results for cond : &&cond&i ";
```

```
proc means data = model&i (where = (bsample > 0)) n min max mean std p5 p95 noprint ;
var beta24 beta72 beta96 ;
output out = result0&i(drop = _TYPE_ _FREQ_) std(beta24 beta72 beta96)= ;
run;
```

```
proc transpose data = result0&i out=stdvec&i (rename = (COL1 = stderr)) ;
run;
```

```

data result&i ;
merge tcond&i stdvec&i ;
by _NAME_ ;
if _NAME_="beta24" then estimate =
  %if &i = 1 %then log(0.978);
  %if &i = 2 %then log(0.920);
  %if &i = 3 %then log(1.150);
  %if &i = 4 %then log(1.299);
  %if &i = 5 %then log(0.879);
  %if &i = 6 %then log(0.959);
  %if &i = 7 %then log(0.99);
  ;
lb = estimate - 1.96 * stderr;
ub = estimate + 1.96 * stderr;
hr = exp(estimate);
lb = exp(lb);
ub = exp(ub);
label hr="Hazard ratio"
      _NAME_=
      time="Time length"
;
length time $20;
if _NAME_="beta24" then time = "2 years";
else if _NAME_ = 'beta72' then time="6 years";
else if _NAME_ = 'beta96' then time="8 years";
run;

proc print data = result&i label noobs;
var time hr stderr lb ub ;
run;

```

```
%end ;
```

```
%mend ;
```

```

%macro run_all(bootstrap = 0, /* Run bootstrap analysis 0 = no, 1 = yes */
  nboot = 200, /* Number of bootstrap samples to run in total */
  bootstart=0, /* Bootstrap sample to start with 0 = original sample, must be less than nboot */
  bootend=200, /* last bootstrap sample to run */
  keep_listing = 1, /* keep output listing for bootstrap samples */
  dataname = linear /* prefix name for dataset to contain individual bootstrap sample results */
);

```

```
/* Seven final models to run */
```

```

%LET cond1 = invasive ^= . ;
%LET cond2 = tothcat = 0 ;
%LET cond3 = tothcat ^= 0 ;
%LET cond4 = menage = 0 ;
%LET cond5 = menage=1 or menage = 2 ;
%LET cond6 = ager = 1 ;
%LET cond7 = ager = 2 or ager = 3 ;
/*****

```

The following macro creates a data set called whi.seeds which contains 201 rows with 7 seeds that are 3,188,736 = 16,608\*2\*96 which are used for generating the simulated data sets used for independent samples. The current program uses at most 100,000 \* 2 \* 96 random seeds for each of the 7 data sets. The values of lstream can be modified, but there is a limit on the number of rows that will be produced in the seeds data set. If lstream is too large there will not be enough seeds in the seeds data set.

This macro takes a long time to run, so it might be best to run it outside the current run\_all macro and place the data set in the whi library so that the table4 macro can find it.

```

*****/
%seedgen(fseed=123,lstream=3188736, seeds_per_line=7 ,nseeds=201);

```

```

%IF &bootstrap = 0 %THEN %DO;
  %LET nboot = 1 ; /* only want to run the analysis on the original data */
  %LET bootstart = 0 ;
  %LET bootend = 0;
%END;

```

```
%initialize_data;
```

```

%DO bsample = &bootstart %TO &bootend ;
  %PUT running bootstrap sample &bsample ;
  footnote "Results for bootstrap sample = &bsample " ;

```

```

%getbootstrap_sample (datain = ipw0 ,
  dataout = ipw,
  bsample = &bsample);

```

```
%IF &bsample > 0 & &keep_listing = 0 %THEN ODS LISTING EXCLUDE ALL ;;
%calculate_weights;
%table4;
```

```
/* Write individual bootstrap sample results to a permanent SAS data set in the whi directory. There
is a separate file for each of the seven conditions */
%DO i = 1 %TO 7;
```

```
DATA whi.&&dataname.&i._&bootstart._&bootend;
SET whi.&&dataname.&i._&bootstart._&bootend MEAN_FOR_model&i ;
RUN;
```

```
PROC DATASETS LIBRARY=work NOLIST ;
DELETE mean_for_model&i;
QUIT;
```

```
%END;
```

```
/* clean up data sets on each run to make sure that each bootstrap sample generates new data sets */
```

```
PROC DATASETS LIBRARY=work NOLIST ;
DELETE ipw dr_bc a b c d e ipw_a main sse2a1 sse2a2 sse2b1 sse2b2 model2a1_1 model2a2_1
model2b1_1 model2b2_1 _idsneeded_
modell1a1 modell1a2 modell1b1 modell1b2 model2a1 model2a2 model2b1 model2b2 ;
```

```
QUIT;
ODS LISTING ;
```

```
%end;
```

```
%mend ;
```

```
options nocenter mprint ;
```

```
%run_all( bootstrap = 1,
nboot = 200,
bootstart = 0,
bootend = 200 ,
keep_listing = 1,
dataname=newhr );
```

```

*****;
*
*
*
*
*
*
*
*
*
*
*
*****;
PROGRAM 3
RISK DIFFERENCE ESTIMATION
This program estimates the risk differences and their 95% confidence intervals.
The analysis uses the permanent SAS dataset, ipwdr_bc, created in Program 1 Data management.
*****;

```

```

%MACRO initialize_data;

%do i = 1 %to 7 ;
  data whi.&&dataname.&i._&bootstart._&bootend;
  run;
%end;

```

```

/*-----
  CLEAN UP DATA FOR IPW ANALYSIS
  -----*/
DATA ipw_a; SET whi.ipwdr_bc (KEEP=
id r invasive invyear death deathdy dyear lastcont /* outcome related */
adhrate estrogen progest estpro openlabel pillfreq1 pillfreq2 /* hormone use */
year dmflag /* year and DM status */
region ethnic educ marital /* demographic factors */
menarche gravid parity brstfed brstdisb agefbir /* reproductive factors */
bcnum brcafrel bkbbonrel /* family history */
age ager diabtrtb smokingb metb bmicxb alcoholb /* established risk factors */
genhelb bodpainb physfunb preventb hopefulb /* physical & mental health */
sermb bisphob multib fruitb vegeb /* medication & diet */
chdb cardiob cancerb bkboneb osteopb /* cvd, cancer, fracture */
boophb nomam2yr /* mammogram & breast exam */
menoage menosymb vasomotorb /* menopause-related */
recency tothcat oc /* hrt and oc use */

```

```

ldiabslf lsmoking lmet lbmicx lalcohol /* established risk factors */
lgenhel lbodpain lphysfun lprevent /* physical & mental health */
lserm lbispho lmulti /* medication & diet */
lchd lcardio lcancer lfracadj losteop /* cvd, cancer, fracture */
looph lbrstexam lmammogrm lbrstbpsy lbrfin lmafin lbrmafin /* mammogram & breast exam */
lmenosym lvasomotor /* menopause-related */
lvagbleed lheavybld lintermit lbleednow lbrsttend lsevtend lbrstchnng; /* side effects */

```

```

/*-----
  RECODE VARIABLES
  -----*/
region2=(region=2); region3=(region=3); region4=(region=4);
ethnic3=(ethnic=3); ethnic4=(ethnic=4); ethnic8=(ethnic=8);
educ1=(educ=1); educ2=(educ=2); educ9=(educ=9);
marital1=(marital=1); marital2=(marital=2); marital3=(marital=3);
ageg2=(55<=age<60); ageg3=(60<=age<65); ageg4=(65<=age<70); ageg5=(70<=age<75);
ageg6=(75<=age<80);
smokingb1=(smokingb=1); smokingb2=(smokingb=2); smokingb9=(smokingb=9);
metb1=(metb=1); metb2=(metb=2); metb3=(metb=3); metb4=(metb=4);
metb9=(metb=9);
bmicxb3=(bmicxb=3); bmicxb4=(bmicxb=4); bmicxb5=(bmicxb=5); bmicxb6=(bmicxb=6);
alcoholb2=(alcoholb=2); alcoholb3=(alcoholb=3); alcoholb4=(alcoholb=4); alcoholb5=(alcoholb=5);
alcoholb6=(alcoholb=6); alcoholb9=(alcoholb=9);
genhelb2=(genhelb=2); genhelb3=(genhelb=3); genhelb4=(genhelb=4); genhelb5=(genhelb=5);
bodpainb2=(bodpainb=2); bodpainb3=(bodpainb=3); bodpainb4=(bodpainb=4); bodpainb5=(bodpainb=5);
bodpainb9=(bodpainb=9);
physfunb2=(physfunb=2); physfunb3=(physfunb=3); physfunb4=(physfunb=4);
preventb1=(preventb=1); preventb2=(preventb=2);
hopefulb1=(hopefulb=1); hopefulb2=(hopefulb=2); hopefulb3=(hopefulb=3); hopefulb5=(hopefulb=5);
hopefulb9=(hopefulb=9);
fruitb1=(fruitb=1); fruitb2=(fruitb=2); fruitb3=(fruitb=3);
vegeb1=(vegeb=1); vegeb2=(vegeb=2); vegeb3=(vegeb=3);
menoage1=(menoage=1); menoage2=(menoage=2); menoage9=(menoage=9);
menosymb1=(menosymb=1); menosymb2=(menosymb=2); menosymb3=(menosymb=3);
recency1=(recency=1); recency2=(recency=2); recency3=(recency=3); recency4=(recency=4);
tothcat1=(tothcat=1); tothcat2=(tothcat=2); tothcat3=(tothcat>=3);

lsmoking1=(lsmoking=1); lsmoking9=(lsmoking=9);
lmet1=(lmet=1); lmet2=(lmet=2); lmet3=(lmet=3); lmet4=(lmet=4);
lmet9=(lmet=9);
lbmicx1=(lbmicx=1); lbmicx3=(lbmicx=3); lbmicx4=(lbmicx=4); lbmicx5=(lbmicx=5);
lbmicx6=(lbmicx=6); lbmicx9=(lbmicx=9);
lalcohol2=(lalcohol=2); lalcohol3=(lalcohol=3); lalcohol4=(lalcohol=4); lalcohol5=(lalcohol=5);
lalcohol6=(lalcohol=6); lalcohol9=(lalcohol=9);
lgenhel2=(lgenhel=2); lgenhel3=(lgenhel=3); lgenhel4=(lgenhel=4); lgenhel5=(lgenhel=5);
lgenhel9=(lgenhel=9);
lbodpain2=(lbodpain=2); lbodpain3=(lbodpain=3); lbodpain4=(lbodpain=4); lbodpain5=(lbodpain=5);

```

```

lbdopain9=(lbdopain=9);
lphysfun2=(lphysfun=2);    lphysfun3=(lphysfun=3);    lphysfun4=(lphysfun=4);
lprevent1=(lprevent=1);    lprevent2=(lprevent=2);    lprevent3=(lprevent=3);    lprevent4=(lprevent=4);
lprevent5=(lprevent=5);    lprevent6=(lprevent=6);
lcardio1=(lcardio=1);      lcardio2=(lcardio=2);      lcardio3=(lcardio=3);
lmenosym1=(lmenosym=1);    lmenosym2=(lmenosym=2);    lmenosym3=(lmenosym=3);
lheavybld1=(lheavybld=1);  lheavybld2=(lheavybld=2);  lheavybld3=(lheavybld=3);  lheavybld4=(lheavybld=4);
lsevtend1=(lsevtend=1);    lsevtend2=(lsevtend>=2);

```

```

menarche1=(menarche=1);    menarche2=(menarche=2);
gravid1=(gravid=1);        gravid2=(gravid=2);        gravid3=(gravid=3);        gravid9=(gravid=9);
parity1=(parity=1);        parity2=(parity in (2,3));  parity3=(parity in (4,5));  parity4=(parity>5);
parity9=(parity=9);
brstfed1=(brstfed=1);      brstfed2=(brstfed=2);
brstdisb1=(brstdisb=1);    brstdisb2=(brstdisb=2);    brstdisb9=(brstdisb=9);
agefbir1=(agefbir=1);      agefbir2=(agefbir=2);      agefbir3=(agefbir=3);      agefbir9=(agefbir=9);
vasomotorb1=(vasomotorb=1);vasomotorb2=(vasomotorb>1);

```

```

lbrfin0=(lbrfin=0);        lbrfin1=(lbrfin=1);        lbrfin2=(lbrfin=2);
lmafin0=(lmafin=0);        lmafin1=(lmafin=1);        lmafin2=(lmafin=2);
lbrmafin0=(lbrmafin=0);    lbrmafin1=(lbrmafin=1);    lbrmafin2=(lbrmafin=2);
lvasomotor1=(lvasomotor=1);lvasomotor2=(lvasomotor>1);

```

```

DROP smokingb metb bmicxb region ethnic educ marital hopefulb genhelb bodpainb fruitb vegeb preventb menosymb
physfunb alcoholb
lsmoking lmet lbmicx lgenhel lbdopain lcardio lphysfun lalcohol lmenosym lheavybld lsevtend lprevent
menarche parity brstfed brstdisb lbrfin lmafin lbrmafin; RUN;

```

```

/*-----
  ASSIGN ADHRATE FOR THOSE W/ MISSING VALUES
  -----*/

```

```

PROC SORT TAGSORT DATA=ipw_a; BY id year; RUN;

```

```

DATA ipw0 _idtmp_ (keep = newid id );
SET ipw_a (RENAME=(adhrate=adhrate_old)); BY id year;

```

```

retain newid ;
if _n_ = 1 then newid = 0 ;
if first.id then do ;
    newid = newid + 1 ;
    output _idtmp_ ;
end;

```

```

adhrate = adhrate_old;
d = RANUNI(1232);

```

```

IF adhrate_old=. THEN DO;
  IF pillfreq1=0 THEN DO;
    IF pillfreq2=0 THEN adhrate = (0.0 + 0.0) / 14;          /* not at all          */
    IF pillfreq2=1 THEN adhrate = (0.0 + 0.5) / 14;          /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (0.0 + 1.5) / 14;          /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (0.0 + 3.5) / 14;          /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (0.0 + 5.5) / 14;          /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (0.0 + 7.0) / 14;          /* every day of the week */
    IF pillfreq2=. THEN adhrate = (0.0 + 7*d) / 14;          /* missing              */
  END;
  IF pillfreq1=1 THEN DO;
    IF pillfreq2=0 THEN adhrate = (0.5 + 0.0) / 14;          /* not at all          */
    IF pillfreq2=1 THEN adhrate = (0.5 + 0.5) / 14;          /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (0.5 + 1.5) / 14;          /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (0.5 + 3.5) / 14;          /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (0.5 + 5.5) / 14;          /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (0.5 + 7.0) / 14;          /* every day of the week */
    IF pillfreq2=. THEN adhrate = (0.5 + 7*d) / 14;          /* missing              */
  END;
  IF pillfreq1=2 THEN DO;
    IF pillfreq2=0 THEN adhrate = (1.5 + 0.0) / 14;          /* not at all          */
    IF pillfreq2=1 THEN adhrate = (1.5 + 0.5) / 14;          /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (1.5 + 1.5) / 14;          /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (1.5 + 3.5) / 14;          /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (1.5 + 5.5) / 14;          /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (1.5 + 7.0) / 14;          /* every day of the week */
    IF pillfreq2=. THEN adhrate = (1.5 + 7*d) / 14;          /* missing              */
  END;
  IF pillfreq1=3 THEN DO;
    IF pillfreq2=0 THEN adhrate = (3.5 + 0.0) / 14;          /* not at all          */
    IF pillfreq2=1 THEN adhrate = (3.5 + 0.5) / 14;          /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (3.5 + 1.5) / 14;          /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (3.5 + 3.5) / 14;          /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (3.5 + 5.5) / 14;          /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (3.5 + 7.0) / 14;          /* every day of the week */
    IF pillfreq2=. THEN adhrate = (3.5 + 7*d) / 14;          /* missing              */
  END;

```

```

END;
IF pillfreq1=4 THEN DO;
  IF pillfreq2=0 THEN adhrate = (5.5 + 0.0) / 14; /* not al all */
  IF pillfreq2=1 THEN adhrate = (5.5 + 0.5) / 14; /* < 1 day per week */
  IF pillfreq2=2 THEN adhrate = (5.5 + 1.5) / 14; /* 1-2 days per week */
  IF pillfreq2=3 THEN adhrate = (5.5 + 3.5) / 14; /* 3-4 days per week */
  IF pillfreq2=4 THEN adhrate = (5.5 + 5.5) / 14; /* 5-6 days per week */
  IF pillfreq2=5 THEN adhrate = (5.5 + 7.0) / 14; /* every day of the week */
  IF pillfreq2=. THEN adhrate = (5.5 + 7*d) / 14; /* missing */
END;
IF pillfreq1=5 THEN DO;
  IF pillfreq2=0 THEN adhrate = (7.0 + 0.0) / 14; /* not al all */
  IF pillfreq2=1 THEN adhrate = (7.0 + 0.5) / 14; /* < 1 day per week */
  IF pillfreq2=2 THEN adhrate = (7.0 + 1.5) / 14; /* 1-2 days per week */
  IF pillfreq2=3 THEN adhrate = (7.0 + 3.5) / 14; /* 3-4 days per week */
  IF pillfreq2=4 THEN adhrate = (7.0 + 5.5) / 14; /* 5-6 days per week */
  IF pillfreq2=5 THEN adhrate = (7.0 + 7.0) / 14; /* every day of the week */
  IF pillfreq2=. THEN adhrate = (7.0 + 7*d) / 14; /* missing */
END;
IF pillfreq1=. THEN DO;
  IF pillfreq2=0 THEN adhrate = (7*d + 0.0) / 14; /* not al all */
  IF pillfreq2=1 THEN adhrate = (7*d + 0.5) / 14; /* < 1 day per week */
  IF pillfreq2=2 THEN adhrate = (7*d + 1.5) / 14; /* 1-2 days per week */
  IF pillfreq2=3 THEN adhrate = (7*d + 3.5) / 14; /* 3-4 days per week */
  IF pillfreq2=4 THEN adhrate = (7*d + 5.5) / 14; /* 5-6 days per week */
  IF pillfreq2=5 THEN adhrate = (7*d + 7.0) / 14; /* every day of the week */
  IF pillfreq2=. THEN adhrate = d;
END;
END;

IF 0<adhrate<=0.01 THEN adhrate=0; /* if taking <1% in a given year, then recode as taking none */
IF adhrate> 1 THEN adhrate=1; /* if taking >100% in a given year, then recode as taking one */

/*-----
  ASSIGN DOSE (INCLUDING NON-STUDY HORMONE USE
-----*/
extra=0; dose_e=0;
IF (estrogen=1 & progest=1) OR estpro=1 OR openlabel=1 THEN DO;
  extra=1; dose_e=RANUNI(2321);
END;

IF r=1 THEN DO;
  IF extra=0 THEN dose=adhrate;
  IF extra=1 THEN dose=adhrate + dose_e;
END;

IF r=0 THEN DO;
  IF extra=0 THEN dose=0;
  IF extra=1 THEN dose=dose_e;
END;

/*-----
  CREATE VARIABLES FOR WEIGHT ESTIMATION
-----*/
IF dose=0 THEN cendose=0; ELSE cendose=1; /* cendose: censoring indicator for those w/ dose=0 */
IF dose>0 THEN logdose=LOG(dose); /* logdose: log-transformed dose for those w/ dose>0 */
ldose = LAG(dose); IF FIRST.id THEN ldose=dose; /* hormone dose in the previous year */
output ipw0;

DROP adhrate_old d;
RUN;

%let _dsid_ = %sysfunc(open(_idtmp_));
%let _numids_ = %sysfunc(attrn(&_dsid_,nobs));
%let _rc_ = %sysfunc(close(&_dsid_));

%if &bootstrap = 1 %then %do;
  data _idholders_ (index = (sample));
  do sample = 1 to &nboot ;
    do newid = 1 to &_numids_;
      output ;
    end;
  end;
run;

proc surveysselect data= _idholders_
  method = urs
  n= &_numids_
  seed = 1232
  out = _idsamples (keep = sample newid numberhits)
  outall
;
strata sample ;
run;

```



```

%end;

%MEND initialize_data;

%MACRO getbootstrap_sample (datain = ,
                           dataout = ,
                           bsample = );

  data _idsneeded_ ;
  set _idsamples (where = (sample = %eval(&bsample))) ;
  run;

  data &dataout ;
  merge &datain _idsneeded_ (keep = newid numberhits ) ;
  if numberhits = 0 then delete ;
  by newid ;

  run;

%MEND getbootstrap_sample;

%MACRO estimate_weights;

/*****
ESTIMATE WEIGHT
*****/
/*-----
PROBABILITY OF HAVING DOSE=0
-----*/
/* model 1a - numerator of sw */
PROC LOGISTIC DATA=ipw(WHERE=( r=0));
MODEL cendose=year year*year
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      ldose ldose*ldose ldose*ldose*ldose / ridging = none ;
freq numberhits;
OUTPUT OUT=modell1a P=num_adh;
RUN;

PROC LOGISTIC DATA=ipw(WHERE = (r = 1));
MODEL cendose=year year*year
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      ldose ldose*ldose ldose*ldose*ldose / ridging = none ;
freq numberhits;
OUTPUT OUT=modell2a P=num_adh;
RUN;

/* model 1b - denominator of sw */
PROC LOGISTIC DATA=ipw(WHERE = (r=0));
MODEL cendose=year year*year
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
      oc brcafrel bkbonrel
      educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
      menarche1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
      region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
      physfunb2--physfunb4 preventb1--preventb2
      chdb cardiob cancerb bkboneb osteopb

      lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
      looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
      lbrstexam lmmogrm lbrstbpsy lbrmafin0--lbrmafin2
      lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
      lchd lcardio1--lcardio3 lcancer lfracadj lsteop
      lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
      lbrsttend lsevtend1--lsevtend2 lbrstchn9
      ldose ldose*ldose ldose*ldose*ldose / ridging = none ;
freq numberhits;
OUTPUT OUT=modell1b P=den_adh;
RUN;

PROC LOGISTIC DATA=ipw(WHERE = (r=1));

```

```

MODEL cendose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
  oc brcafrel bkbnonrel
  educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
  menarche1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
  region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
  physfunb2--physfunb4 preventb1--preventb2
  chdb cardiob cancerb bkboneb osteopb

  lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
  looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
  lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
  lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
  lchd lcardio1--lcardio3 lcancer lfracadj lsteop
  lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
  lbrsttend lsevtend1--lsevtend2 lbrstchnng
  ldose ldose*ldose ldose*ldose*ldose / ridging = none ;

freq numberhits;
OUTPUT OUT=model1b2 P=den_adh;
RUN;

/*-----
  PROBABILITY OF HAVING A PARTICULAR DOSE AMONG THOSE W/ DOSE>0
  -----*/

/* model 2a - numerator of sw */
PROC GLM DATA=ipw(WHERE = (cendose=1 AND r=0))
  ORDER=FREQ OUTSTAT=sse2a1(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafrel
  tothcat1--tothcat3
  menarche1--menarche2
  marital1--marital3 menoage1--menoage9
  ldose ldose*ldose ldose*ldose*ldose / solution;

freq numberhits;
OUTPUT OUT=model2a1_1 P=pred;
RUN;
QUIT;

PROC GLM DATA=ipw(WHERE = (cendose=1 AND r=1))
  ORDER=FREQ OUTSTAT=sse2a2(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafrel
  tothcat1--tothcat3
  menarche1--menarche2
  marital1--marital3 menoage1--menoage9
  ldose ldose*ldose ldose*ldose*ldose / solution;

freq numberhits;
OUTPUT OUT=model2a2_1 P=pred;
RUN;
QUIT;

/* MODEL 2b - denominator of sw */
PROC GLM DATA=ipw(WHERE = (cendose=1 AND r=0))
  ORDER=FREQ OUTSTAT=sse2b1(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
  oc brcafrel bkbnonrel
  educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
  menarche1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
  region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
  physfunb2--physfunb4 preventb1--preventb2
  chdb cardiob cancerb bkboneb osteopb

  lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
  looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
  lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
  lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
  lchd lcardio1--lcardio3 lcancer lfracadj lsteop
  lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
  lbrsttend lsevtend1--lsevtend2 lbrstchnng
  ldose ldose*ldose ldose*ldose*ldose / solution;

freq numberhits;
OUTPUT OUT=model2b1_1 P=pred R=residual;
RUN;
QUIT;

PROC GLM DATA=ipw(WHERE = (cendose=1 AND r=1))
  ORDER=FREQ OUTSTAT=sse2b2(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8

```

```
lbmicxb3--lbmicxb6 metb1--metb9 smokngb1--smokngb9 lalcoholb2--lalcoholb9
parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
oc brcafrel bkbbonrel
educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
menarchel1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoagel1--menoage9
physfunb2--physfunb4 preventb1--preventb2
chdb cardiob cancerb bkbboneb osteopb
```

```
lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
lchd lcardio1--lcardio3 lcancer lfracadj lsteop
lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
lbrsttend lsevtend1--lsevtend2 lbrstchnge
ldose ldose*ldose ldose*ldose*ldose / solution;
```

```
freq numberhits;
```

```
OUTPUT OUT=model2b2_1 P=pred R=residual;
```

```
RUN;
```

```
QUIT;
```

```
DATA sse2a1;
```

```
SET sse2a1;
```

```
rootmse_n=SQRT(ss/df);
```

```
RUN;
```

```
DATA sse2a2;
```

```
SET sse2a2;
```

```
rootmse_n=SQRT(ss/df);
```

```
RUN;
```

```
DATA sse2b1;
```

```
SET sse2b1;
```

```
rootmse_n=SQRT(ss/df);
```

```
RUN;
```

```
DATA sse2b2;
```

```
SET sse2b2;
```

```
rootmse_n=SQRT(ss/df);
```

```
RUN;
```

```
PROC SQL;
```

```
CREATE TABLE model2a1 AS
```

```
SELECT O.*, L.rootmse_n
```

```
FROM model2a1_1 O, sse2a1 L;
```

```
QUIT;
```

```
PROC SQL;
```

```
CREATE TABLE model2a2 AS
```

```
SELECT O.*, L.rootmse_n
```

```
FROM model2a2_1 O, sse2a2 L;
```

```
QUIT;
```

```
PROC SQL;
```

```
CREATE TABLE model2b1 AS
```

```
SELECT O.*, L.rootmse_n
```

```
FROM model2b1_1 O, sse2b1 L;
```

```
QUIT;
```

```
PROC SQL;
```

```
CREATE TABLE model2b2 AS
```

```
SELECT O.*, L.rootmse_n
```

```
FROM model2b2_1 O, sse2b2 L;
```

```
QUIT;
```

```
DATA model2a1;
```

```
SET model2a1;
```

```
num_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);
```

```
FORMAT num_adhrate 12.10;
```

```
RUN;
```

```
DATA model2a2;
```

```
SET model2a2;
```

```
num_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);
```

```
FORMAT num_adhrate 12.10;
```

```
RUN;
```

```
DATA model2b1;
```

```
SET model2b1;
```

```
den_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);
```

```
FORMAT den_adhrate 12.10;
```

```
RUN;
```

```
DATA model2b2;
```

```
SET model2b2;
```

```
den_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);
```

```
FORMAT den_adhrate 12.10;
```

```
RUN;
```

```

/*-----
ESTIMATE STABILIZED WEIGHT
-----*/
DATA ipw;
SET ipw (KEEP= numberhits id r adhrate dose dose_e extra cendose year invasive
  invyear invady death dyear deathdy lastcont ager menoage dmflag age ageg2--ageg6
  ethnic3--ethnic8 bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2 oc brcafrel tothcat
  tothcat1--tothcat3 menarchel1--menarche2 marital1--marital3 menoage1--menoage9);
RUN;

DATA modella1;
SET modella1 (KEEP=id year num_adh);
RUN;

DATA modella2;
SET modella2 (KEEP=id year num_adh);
RUN;

DATA modellb1;
SET modellb1 (KEEP=id year den_adh);
RUN;

DATA modellb2;
SET modellb2 (KEEP=id year den_adh);
RUN;

DATA model2a1;
SET model2a1 (KEEP=id year num_adhrate);
RUN;

DATA model2a2;
SET model2a2 (KEEP=id year num_adhrate);
RUN;

DATA model2b1;
SET model2b1 (KEEP=id year den_adhrate);
RUN;

DATA model2b2;
SET model2b2 (KEEP=id year den_adhrate);
RUN;

PROC SORT TAGSORT DATA=ipw;
BY id year;
RUN;

PROC SORT TAGSORT DATA=modella1;
BY id year;
RUN;

PROC SORT TAGSORT DATA=modella2;
BY id year;
RUN;

PROC SORT TAGSORT DATA=modellb1;
BY id year;
RUN;

PROC SORT TAGSORT DATA=modellb2;
BY id year;
RUN;

PROC SORT TAGSORT DATA=model2a1;
BY id year;
RUN;

PROC SORT TAGSORT DATA=model2a2;
BY id year;
RUN;

PROC SORT TAGSORT DATA=model2b1;
BY id year;
RUN;

PROC SORT TAGSORT DATA=model2b2;
BY id year;
RUN;

DATA main;
MERGE ipw modella1 modella2 modellb1 modellb2 model2a1 model2a2 model2b1 model2b2;
BY id year;
IF FIRST.id THEN stabw=1;
RETAIN stabw;

w2 =      num_adh      /      den_adh;
w3 = (1 - num_adh)    / (1 - den_adh);
w4 =      num_adhrate /      den_adhrate;

IF cendose=0 THEN stabw = stabw * w2;
IF cendose=1 THEN stabw = stabw * w3 * w4;

```

DROP w2-w4; RUN;

/\*-----  
EXPAND DATASET INTO MONTH PER OBSERVATION  
-----\*/

/\* create 'maxyear' to be used for do loop \*/  
PROC SQL;  
CREATE TABLE a AS  
SELECT id, MAX(year) AS maxyear  
FROM main GROUP BY id;  
QUIT;

PROC SQL;  
CREATE TABLE b AS  
SELECT O.\*, L.maxyear  
FROM main O, a L  
WHERE O.id=L.id;  
QUIT;

/\* create month of invasive breast cancer (invamo) \*/  
DATA c;  
SET b;  
invamo = CEIL(invady/30.4375);  
deathmo = CEIL(deathdy/30.4375);

mm=0;  
RETAIN mm;  
DO y=1 UNTIL (year=maxyear);  
DO m=1 TO 12;  
mm=mm+1;  
OUTPUT;  
END;  
END;  
DROP y m;  
RUN;

/\* create months since randomization (month) & delete observation after bc or death occurred, or last follow-up month \*/  
DATA d; SET c;  
month=mm+12\*(year-1);

maxmo = maxyear\*12;  
lmo = CEIL(lastcont/30.4375);  
lyear = CEIL(lastcont/365.25);  
lastmo = MIN(lmo,maxmo); /\* lmo: last month based on lastcont; maxmo: last month based on data \*/  
lastyear = MIN(lyear,maxyear); /\* lyear: last year based on lastcont; maxyear: last year based on data \*/

IF invamo NE . AND month > invamo THEN DELETE;  
IF deathmo NE . AND month > deathmo THEN DELETE;  
IF invamo = . AND deathmo = . AND month > lastmo THEN DELETE;

IF invamo NE . AND invamo = month THEN invasive=1;  
ELSE invasive=0;  
IF deathmo NE . AND deathmo = month THEN death=1;  
ELSE death=0;

DROP mm;  
RUN;

/\*-----  
CALCULATE CUMULATIVE USE  
-----\*/

DATA p;  
SET whi.baseline\_bc(KEEP=id);  
p=RANUNI(99999); RUN;

DATA e;  
MERGE p d (in = a);  
BY id;  
if a ;

RUN;

PROC SORT TAGSORT DATA=e;  
BY id month;  
RUN;

DATA dr\_bc(KEEP= numberhits cuma stabw id r month year invasive ager menoa9e dmflag  
ageg2--ageg6 ethnic3--ethnic8 bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9  
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2 oc brcafre1 tothcat tothcat1--tothcat3  
menarche1--menarche2 marital1--marital3 menoa9e1--menoa9e9);

```

SET e;
BY id month;

IF FIRST.id THEN cuma = 0;
RETAIN cuma;

/* during the non-last available year when bc or death did not occur, dose is equally distributed among 12 months */
IF (year NE lastyear) & (year NE invyear) & (year NE dyear) THEN DO;
  cuma = cuma + (dose / 12);
END;

/* during the last available year when bc or death did not occur, dose is equally distributed among available months */
IF (year=lastyear) & (year NE invyear) & (year NE dyear) THEN DO;
  d0 = (lastmo-((lastyear-1)*12));
  cuma = cuma + (dose / d0);
END;

/* during the year when death occurred, dose is randomly distributed among months until death */
IF year = dyear THEN DO;
  d1 = (deathmo-((dyear-1)*12));
  cuma = cuma + (dose / d1);
END;

/* during the year when bc occurred & invyear ne lastyear, dose is randomly assigned to distributed among avail. months */
IF (year = invyear) & (invyear NE lastyear) & (year NE dyear) THEN DO;
  d2 = invamo-((invyear-1)*12);
  d3 = 13 - d2;
  IF 0/d3 <= p <= 1/d3 THEN cuma = cuma + (dose / d2);
  IF 1/d3 < p <= 2/d3 THEN cuma = cuma + (dose / (d2+1));
  IF 2/d3 < p <= 3/d3 THEN cuma = cuma + (dose / (d2+2));
  IF 3/d3 < p <= 4/d3 THEN cuma = cuma + (dose / (d2+3));
  IF 4/d3 < p <= 5/d3 THEN cuma = cuma + (dose / (d2+4));
  IF 5/d3 < p <= 6/d3 THEN cuma = cuma + (dose / (d2+5));
  IF 6/d3 < p <= 7/d3 THEN cuma = cuma + (dose / (d2+6));
  IF 7/d3 < p <= 8/d3 THEN cuma = cuma + (dose / (d2+7));
  IF 8/d3 < p <= 9/d3 THEN cuma = cuma + (dose / (d2+8));
  IF 9/d3 < p <= 10/d3 THEN cuma = cuma + (dose / (d2+9));
  IF 10/d3 < p <= 11/d3 THEN cuma = cuma + (dose / (d2+10));
  IF 11/d3 < p <= 12/d3 THEN cuma = cuma + (dose / (d2+11));
END;

/* during the year when bc occurred & invyear=lastyear, dose is randomly assigned to distributed among avail. months */
IF (year = invyear) & (invyear=lastyear) & (year NE dyear) THEN DO;
  d4 = invamo-((invyear-1)*12);
  d5 = lastmo-((lastyear-1)*12);
  d6 = (d5-d4) + 1;

  IF 0/d6 <= p <= 1/d6 THEN cuma = cuma + (dose / d4);
  IF 1/d6 < p <= 2/d6 THEN cuma = cuma + (dose / (d4+1));
  IF 2/d6 < p <= 3/d6 THEN cuma = cuma + (dose / (d4+2));
  IF 3/d6 < p <= 4/d6 THEN cuma = cuma + (dose / (d4+3));
  IF 4/d6 < p <= 5/d6 THEN cuma = cuma + (dose / (d4+4));
  IF 5/d6 < p <= 6/d6 THEN cuma = cuma + (dose / (d4+5));
  IF 6/d6 < p <= 7/d6 THEN cuma = cuma + (dose / (d4+6));
  IF 7/d6 < p <= 8/d6 THEN cuma = cuma + (dose / (d4+7));
  IF 8/d6 < p <= 9/d6 THEN cuma = cuma + (dose / (d4+8));
  IF 9/d6 < p <= 10/d6 THEN cuma = cuma + (dose / (d4+9));
  IF 10/d6 < p <= 11/d6 THEN cuma = cuma + (dose / (d4+10));
  IF 11/d6 < p <= 12/d6 THEN cuma = cuma + (dose / (d4+11));
END;
RUN;

```

```
%MEND estimate_weights;
```

```
%MACRO numargs(arg);
```

```

  %let n = 1;
  %if &arg^= %then %do;
    %do %until (%scan(&arg,%eval(&n),%str( ))=%str());

      %let word = %scan(&arg,&n);
      %let n = %eval(&n+1);

    %end;

  %end;
  %eval(&n-1) /* there is no ; here since it will be used as %let a = %numargs(&b) ;
              and the ; is included at the end of this line */

```

```
%MEND numargs;
```

```
%MACRO risk_diff;
```

```
DATA r;
```

```

SET DR_BC;
%rcspline(month,6,48,90);
use1      = cuma      * (ager=1);
use2      = cuma      * (ager in (2,3));

interact1 = use1 * month;
interact2 = use1 * month1;
interact3 = use2 * month;
interact4 = use2 * month1;

a = (tothcat ne 0) * month;
b = (tothcat ne 0) * month1;
RUN;

/* PROC LOGISTIC generated the same curve as PROC GENMOD */
PROC LOGISTIC DESCENDING DATA=R;
ODS OUTPUT ParameterEstimates=PE;
MODEL Invasive = use1-use2
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      interact1-interact4 a b ;
FREQ numberhits ;
WEIGHT stabw;
RUN;

/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
  SELECT estimate FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' '
  FROM pe;
QUIT;

PROC SQL NOPRINT;
  SELECT variable INTO:model SEPARATED BY ' '
  FROM pe;
QUIT;

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;

PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' '
FROM nobs;
QUIT;

DATA r1 (KEEP=month s0 s1 numberhits);
SET r( where = ( ager=1 AND month = 1));

ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

intercept = 1;
s0        = 1;
s1        = 1;

DO month=1 TO 96;
  %rcspline(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  do i=1 to DIM(var);
    use1      = 0;
    use2      = 0;
    interact1 = 0;
    interact2 = 0;
    interact3 = 0;
    interact4 = 0;

    xbeta0    = xbeta0 + coef[i] * var[i];

    use1      = (month/12)          * (ager=1);
    use2      = (month/12)          * (ager IN (2,3));
  end;
end;

```

```

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

xbetal = xbetal + coef[i] * var[i];
end;

p0 = 1 / (1 + EXP(-xbeta0));
p1 = 1 / (1 + EXP(-xbetal));
s0 = s0 * (1 - p0);
s1 = s1 * (1 - p1);
if month IN (24,72,96) then OUTPUT;
END; RUN;

PROC MEANS DATA=r1 MEAN NOPRINT;
WHERE month NE .;
freq numberhits ;
VAR s0 s1;
OUTPUT OUT=MEAN_FOR_KM1(DROP=_TYPE_ _FREQ_) MEAN(s0 s1)=;
CLASS month;
RUN;

DATA mean_for_model6;
SET MEAN_FOR_KM1;
LABEL years = 'Years since randomization'
month = 'Months since randomization'
s0 = 'Never treated'
s1 = 'Always treated';
years=month/12;
bsample = &bsample ;
RUN;

/*****/

DATA r;
SET dr_bc;
%rcspline(month,6,48,90);
use1 = cuma * (ager=1);
use2 = cuma * (ager IN (2,3));

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
RUN;

PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use2
month month1
dmflag ageg2--ageg6 ethnic3--ethnic8
bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
oc brcafrel
tothcat1--tothcat3
menarchel--menarche2
marital1--marital3 menoage1--menoage9
interact1-interact4 a b ;
freq numberhits ;
WEIGHT stabw;
RUN;

/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;

PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nobs;
QUIT;

```



```

DATA r1 (KEEP=month s0 s1 numberhits);
SET r( WHERE = ( month = 1 and  AGER IN (2,3)));

ARRAY var{&NVAR}    &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

intercept = 1;
s0        = 1;
s1        = 1;

DO month=1 TO 96;
  %rcspline(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  DO i=1 TO DIM(var);
    use1      = 0;
    use2      = 0;
    interact1 = 0;
    interact2 = 0;
    interact3 = 0;
    interact4 = 0;

    xbeta0    = xbeta0 + coef[i] * var[i];

    use1      = (month/12)          * (ager=1);
    use2      = (month/12)          * (ager IN (2,3));
    interact1 = use1 * month;
    interact2 = use1 * month1;

    interact3 = use2 * month;
    interact4 = use2 * month1;

    xbeta1    = xbeta1 + coef[i] * var[i];
  end;

  p0 = 1 / (1 + EXP(-xbeta0));
  p1 = 1 / (1 + EXP(-xbeta1));
  s0 = s0 * (1 - p0);
  s1 = s1 * (1 - p1);
  IF month IN (24,72,96) THEN OUTPUT;
END;
RUN;

PROC MEANS DATA=r1 MEAN NOPRINT;
WHERE month NE .;
VAR s0 s1;
freq numberhits ;
OUTPUT OUT=mean_for_kml(DROP=_TYPE_ _FREQ_) MEAN(s0 s1)=;
CLASS MONTH;
RUN;

DATA mean_for_model7;
SET mean_for_kml;
LABEL years = 'Years since randomization'
      month = 'Months since randomization'
      s0    = 'Never treated'
      s1    = 'Always treated';
years=month/12;
bsample = &bsample ;
RUN;
/*****/

DATA r;
SET dr_bc;
%rcspline(month,6,48,90);
use1      = cuma          * (menoage=0);
use2      = cuma          * (menoage IN (1,2));
use3      = cuma          * (menoage=9);

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

```

```
interact5 = use3 * month;
interact6 = use3 * month1;
```

```
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
```

```
run;
```

```
/* PROC LOGISTIC generated the same curve as PROC GENMOD */
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use3
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      interact1-interact6 a b ;
freq numberhits ;
WEIGHT stabw;
RUN;
```

```
/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;
```

```
PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM PE;
QUIT;
```

```
/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=NOBS(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nob;
QUIT;
```

```
DATA R1 (KEEP=month s0 s1 numberhits);
SET r (WHERE = (month=1 and menoage=0));
```

```
ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);
```

```
intercept = 1;
s0         = 1;
s1         = 1;
```

```
DO month=1 TO 96;
  %rcspline(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;
```

```
DO i=1 TO DIM(var);
  use1 = 0; /* never treated */
  use2 = 0;
  use3 = 0;

  interact1 = 0;
  interact2 = 0;
  interact3 = 0;
  interact4 = 0;
  interact5 = 0;
  interact6 = 0;

  xbeta0 = xbeta0 + coef[i] * var[i];

  use1 = (month/12) * (menoage=0); /* always treated */
  use2 = (month/12) * (menoage IN (1,2));
  use3 = (month/12) * (menoage=9);
```

```

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

interact5 = use3 * month;
interact6 = use3 * month1;

xbetal = xbetal + coef[i] * var[i];
end;

p0 = 1 / (1 + EXP(-xbeta0));
p1 = 1 / (1 + EXP(-xbetal));
s0 = s0 * (1 - p0);
s1 = s1 * (1 - p1);
IF month in (24,72,96) THEN OUTPUT;
END;
RUN;

PROC MEANS DATA=r1 MEAN NOPRINT;
WHERE month NE .;
VAR s0 s1;
freq numberhits ;
OUTPUT OUT=mean_for_kml(DROP=_TYPE_ _FREQ_) MEAN(S0 S1)=;
CLASS month;
RUN;

DATA mean_for_model4;
SET mean_for_kml;
LABEL years = 'Years since randomization'
      month = 'Months since randomization'
      s0 = 'Never treated'
      s1 = 'Always treated';
years=month/12;
bsample = &bsample ;
RUN;

/*****/

DATA r;
SET dr_bc;
%rcspline(month,6,48,90);
use1 = cuma * (menoage=0);
use2 = cuma * (menoage IN (1,2));
use3 = cuma * (menoage=9);

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

interact5 = use3 * month;
interact6 = use3 * month1;

a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
RUN;

/* PROC LOGISTIC generated the same curve as PROC GENMOD */
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use3
      month month1
      dmflag age2--age6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafre1
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      interact1-interact6 a b ;

freq numberhits ;
WEIGHT stabw;
RUN;

/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;

```

```

PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nobs;
QUIT;

DATA r1 (KEEP=month s0 s1 numberhits);
SET r ;
WHERE menoage IN (1,2);
if month = 1;

ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

intercept = 1;
s0 = 1;
s1 = 1;

DO month=1 TO 96;
  %rcspline(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  DO i=1 TO DIM(var);
    use1 = 0; /* never treated */
    use2 = 0;
    use3 = 0;

    interact1 = 0;
    interact2 = 0;
    interact3 = 0;
    interact4 = 0;
    interact5 = 0;
    interact6 = 0;

    xbeta0 = xbeta0 + coef[i] * var[i];

    use1 = (month/12) * (menoage=0); /* always treated */
    use2 = (month/12) * (menoage IN (1,2));
    use3 = (month/12) * (menoage=9);

    interact1 = use1 * month;
    interact2 = use1 * month1;

    interact3 = use2 * month;
    interact4 = use2 * month1;

    interact5 = use3 * month;
    interact6 = use3 * month1;

    xbeta1 = xbeta1 + coef[i] * var[i];
  END;

  p0 = 1 / (1 + EXP(-xbeta0));
  p1 = 1 / (1 + EXP(-xbeta1));
  s0 = s0 * (1 - p0); /* never treated */
  s1 = s1 * (1 - p1); /* always treated */
  IF month in (24,72,96) THEN OUTPUT;
END;
RUN;

PROC MEANS DATA=r1 MEAN NOPRINT; WHERE MONTH NE .;
VAR s0 s1;
FREQ numberhits ;
OUTPUT OUT=mean_for_kml(DROP=_TYPE_ _FREQ_) MEAN(s0 s1)=;
CLASS MONTH; RUN;

DATA mean_for_model5;
SET mean_for_kml;
LABEL years = 'Years since randomization'

```

```

    month = 'Months since randomization'
    s0    = 'Never treated'
    s1    = 'Always treated';
years=month/12;
bsample = &bsample ;
RUN;

/*****/

DATA r;
SET dr_bc;
%rcspline(month,6,48,90);
use1      = cuma      * (tothcat=0) ;
use2      = cuma      * (tothcat NE 0);

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;

RUN;

/* PROC LOGISTIC generated the same curve as PROC GENMOD */
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use2
              month month1
              dmflag age2--age6 ethnic3--ethnic8
              bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
              parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
              oc bracafrel
              tothcat1--tothcat3
              menarche1--menarche2
              marital1--marital3 menoage1--menoage9
              interact1-interact4 a b ;

FREQ numberhits ;
WEIGHT stabw;
RUN;

/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;

PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR DF;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nobs;
QUIT;

DATA r1 (KEEP=month s0 s1 numberhits );
SET r;
WHERE tothcat=0 AND month = 1;

ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

intercept = 1;
s0        = 1;
s1        = 1;

DO month=1 TO 96;
  %rcspline(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;

  DO i=1 TO DIM(var);

```

```

use1      = 0;                               /* never treated */
use2      = 0;

interact1 = 0;
interact2 = 0;
interact3 = 0;
interact4 = 0;

xbeta0    = xbeta0 + coef[i] * var[i];

use1      = (month/12)                       * (tothcat=0) ;           /* always treated */
use2      = (month/12)                       * (tothcat NE 0);

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

xbeta1    = xbeta1 + coef[i] * var[i];
END;

p0 = 1 / (1 + EXP(-xbeta0));
p1 = 1 / (1 + EXP(-xbeta1));
s0 = s0 * (1 - p0);                          /* never treated */
s1 = s1 * (1 - p1);                          /* always treated */
IF month in (24,72,96) THEN OUTPUT;
END;
RUN;

PROC MEANS DATA=r1 MEAN NOPRINT;
WHERE month NE .;
VAR s0 s1;
freq numberhits ;
OUTPUT OUT=mean_for_kml(DROP=_TYPE_ _FREQ_) MEAN(s0 s1)=;
CLASS month;
RUN;

DATA mean_for_model2;
SET mean_for_kml;
LABEL years = 'Years since randomization'
      month = 'Months since randomization'
      s0    = 'Never treated'
      s1    = 'Always treated';
years=month/12;
bsample = &bsample ;
RUN;

/*****/

DATA r;
SET dr_bc;
%rcspline(month,6,48,90);
use1      = cuma      * (tothcat=0) ;
use2      = cuma      * (tothcat NE 0);

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;

RUN;

/* PROC LOGISTIC generated the same curve as PROC GENMOD */
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=PE;
MODEL invasive = use1-use2
      month month1
      dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
      parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
      oc brcafrel
      tothcat1--tothcat3
      menarche1--menarche2
      marital1--marital3 menoage1--menoage9
      interact1-interact4 a b ;
freq numberhits ;
WEIGHT stabw;
RUN;

/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;

```

```

QUIT;

PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;

/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=(_STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nobs;
QUIT;

DATA r1 (KEEP=month s0 s1 numberhits);
SET r;
WHERE month = 1 and tothcat NE 0;

ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);

intercept = 1;
s0 = 1;
s1 = 1;

DO month=1 TO 96;
%rcspline(month,6,48,90);
xbeta0 = 0;
xbetal = 0;
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;

DO i=1 TO DIM(var);
use1 = 0; /* never treated */
use2 = 0;

interact1 = 0;
interact2 = 0;
interact3 = 0;
interact4 = 0;

xbeta0 = xbeta0 + coef[i] * var[i];

use1 = (month/12) * (tothcat=0) ; /* always treated */
use2 = (month/12) * (tothcat NE 0);

interact1 = use1 * month;
interact2 = use1 * month1;

interact3 = use2 * month;
interact4 = use2 * month1;

xbetal = xbetal + coef[i] * var[i];
END;

p0 = 1 / (1 + EXP(-xbeta0));
p1 = 1 / (1 + EXP(-xbetal));
s0 = s0 * (1 - p0); /* never treated */
s1 = s1 * (1 - p1); /* always treated */
IF month IN (24,72,96) THEN OUTPUT;
END;
RUN;

PROC MEANS DATA=r1 MEAN NOPRINT;
WHERE month NE .;
VAR s0 s1;
freq numberhits ;
OUTPUT OUT=mean_for_kml(DROP=_TYPE_ _FREQ_) MEAN(s0 s1)=;
CLASS month;
RUN;

DATA mean_for_model3;
SET mean_for_kml;
LABEL years = 'Years since randomization'
month = 'Months since randomization'
s0 = 'Never treated'
s1 = 'Always treated';
years=month/12;
bsample = &bsample ;

```

RUN;

/\* \*\*\*\* \*/

```
DATA r;
SET dr_bc;
%rcspline(month,6,48,90);
cuma2 = cuma * cuma;
interact1 = cuma * month;
interact2 = cuma * month1;
```

```
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
RUN;
```

```
/* PROC LOGISTIC generated the same curve as PROC GENMOD */
PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=pe;
MODEL invasive = cuma
              month month1
              dmflag ageg2--ageg6 ethnic3--ethnic8
              bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
              parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
              oc brcafrel
              tothcat1--tothcat3
              menarche1--menarche2
              marital1--marital3 menoage1--menoage9
              interact1-interact2 a b ;
FREQ numberhits ;
WEIGHT stabw;
RUN;
```

```
/* create a macro variable IBC_ESTIMATE that carries the values of the coefficients */
PROC SQL NOPRINT;
SELECT ESTIMATE FORMAT=16.12 INTO:IBC_ESTIMATE SEPARATED BY ' ' FROM pe;
QUIT;
```

```
PROC SQL NOPRINT;
SELECT variable INTO:model SEPARATED BY ' ' FROM pe;
QUIT;
```

```
/* create a macro variable NVAR that carries the total number of baseline variables in the weighted model (+ intercept) */
PROC MEANS SUM NOPRINT DATA=pe;
VAR df;
OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N"));
RUN;
PROC SQL NOPRINT;
SELECT DF INTO:NVAR SEPARATED BY ' ' FROM nobs;
QUIT;
```

```
DATA r1 (KEEP=month s0 s1 numberhits);
SET r;
WHERE month = 1;
```

```
ARRAY var{&NVAR} &MODEL;
ARRAY coef{&NVAR} (&IBC_ESTIMATE);
```

```
intercept = 1;
s0 = 1;
s1 = 1;
```

```
DO month=1 TO 96;
  %rcspline(month,6,48,90);
  xbeta0 = 0;
  xbeta1 = 0;
  a = (tothcat NE 0) * month;
  b = (tothcat NE 0) * month1;
```

```
DO i=1 TO DIM(var);
  cuma = 0; /* never treated */

  interact1 = 0;
  interact2 = 0;

  xbeta0 = xbeta0 + coef[i] * var[i];

  cuma = (month/12); /* always treated */
```



```

interact1 = (month/12)          * month;
interact2 = (month/12)          * month1;

    xbeta1    = xbeta1 + coef[i] * var[i];
END;

p0 = 1 / (1 + EXP(-xbeta0));
p1 = 1 / (1 + EXP(-xbeta1));
s0 = s0 * (1 - p0);
s1 = s1 * (1 - p1);
IF month IN (24,72,96) THEN OUTPUT;
/* never treated */
/* always treated */
END;
RUN;

PROC MEANS DATA=r1 MEAN NOPRINT;
WHERE month NE .;
VAR s0 s1;
FREQ numberhits ;

OUTPUT OUT=mean_for_kml(DROP=_TYPE_ _FREQ_) MEAN(s0 s1)=;
CLASS month;
RUN;

DATA mean_for_model1;
SET mean_for_kml;
LABEL years = 'Years since randomization'
      month = 'Months since randomization'
      s0     = 'Never treated'
      s1     = 'Always treated';
years=month/12;
bsample = &bsample ;
RUN;

PROC DATASETS LIBRARY = work NOLIST ;
DELETE p pe r r1 mean_for_kml ;
QUIT;

%MEND risk_diff;

%MACRO rcspline(x,knot1,knot2,knot3,knot4,knot5,knot6,knot7,knot8,knot9,knot10, norm=2);
%LOCAL j v7 k tk tk1 t k1 k2;
%LET v7=&x; %IF %length(&v7)=8 %THEN %LET v7=%SUBSTR(&v7,1,7);
%*get no. knots, last knot, next to last knot;
%DO k=1 %TO 10;
%IF %QUOTE(&k)=% THEN %GOTO nomorek;
%END;
%LET k=11;
%nomorek: %LET k=%EVAL(&k-1); %LET k1=%EVAL(&k-1); %LET k2=%EVAL(&k-2);
%IF &k<3 %THEN %PUT ERROR: <3 knots given. no spline variables CREATED.;
%ELSE %DO;
%LET tk=&knot&k;
%LET tk1=&knot&k1;
DROP _kd_ ; _kd_=
%IF &norm=0 %THEN 1;
%ELSE %IF &norm=1 %THEN &tk - &tk1;
%ELSE (&tk - &knot1)**.6666666666666666 ;
%DO j=1 %TO &k2;
%LET t=&knot&j;
&v7&j=max((&x-&t)/_kd_,0)**3+((&tk1-&t)*max((&x-&tk)/_kd_,0)**3
-(&tk-&t)*max((&x-&tk1)/_kd_,0)**3)/(&tk-&tk1)%STR(;;);
%END;
%END;
%MEND rcspline;

%MACRO run_all(bootstrap = 0, /* Run bootstrap analysis 0 = no, 1 = yes */
              nboot = 200, /* Number of bootstrap samples to run in total */
              bootstart=0, /* Bootstrap sample to start with 0 = original sample, must be less than nboot */
              bootend=200, /* last bootstrap sample to run */
              keep_listing = 1, /* keep output listing for bootstrap samples */
              dataname = linear /* prefix name for dataset to contain individual bootstrap sample results */
              );

/* Seven final models to run */
%LET cond1 = invasive ^= . ;
%LET cond2 = tothcat = 0 ;
%LET cond3 = tothcat ^= 0 ;
%LET cond4 = menoage = 0 ;
%LET cond5 = menoage =1 or menoage = 2 ;
%LET cond6 = ager = 1 ;
%LET cond7 = ager = 2 or ager = 3 ;

%IF &bootstrap = 0 %THEN %DO;
%LET nboot = 0 ; /* only want to run the analysis on the original data */

```

```

%LET bootstart = 0 ;
%LET bootend = 0;
%END;

%initialize_data;

%DO bsample = &bootstart %TO &bootend ;
  %PUT running bootstrap sample &bsample ;
  footnote "Results for bootstrap sample = &bsample " ;

  %IF &bsample > 0 %THEN %DO;
    %getbootstrap_sample (datain = ipw0 ,
                        dataout = ipw,
                        bsample = &bsample);

  %END;
%ELSE %DO;
  DATA ipw;
  SET ipw0;
  numberhits = 1 ;
  RUN;
%END;

%IF &bsample > 0 & &keep_listing = 0 %THEN ODS LISTING EXCLUDE ALL ;;
%estimate_weights;
%risk_diff;

/* Write individual bootstrap sample results to a permanent SAS data set in the whi directory. There
is a separate file for each of the seven conditions */
%DO i = 1 %TO 7;

  DATA whi.&&dataname.&i._&bootstart._&bootend;
  SET whi.&&dataname.&i._&bootstart._&bootend MEAN_FOR_model&i ;
  RUN;

  PROC DATASETS LIBRARY=work NOLIST ;
  DELETE mean_for_model&i;
  QUIT;

  %END;
/* clean up data sets on each run to make sure that each bootstrap sample generates new data sets */
PROC DATASETS LIBRARY=work NOLIST ;
DELETE ipw dr_bc a b c d e ipw_a main sse2a1 sse2a2 sse2b1 sse2b2 model2a1_1 model2a2_1
      model2b1_1 model2b2_1 _idsneeded_
      modell1a1 modell1a2 modell1b1 modell1b2 model2a1 model2a2 model2b1 model2b2 ;
QUIT;
ODS LISTING ;
%end;

%MEND run_all;

%MACRO results;

%LET cond1 = Overall ;
%LET cond2 = No prior hormone use ;
%LET cond3 = Prior hormone use ;
%LET cond4 = Years since menopause < 10 ;
%LET cond5 = Years since menopause > 10 ;
%LET cond6 = Age <= 60 ;
%LET cond7 = Age > 60 ;

DATA table4rd ;
RUN;

%DO i = 1 %TO 7;

  /* NOTE: default is 200 bootstraps, therefore the datanames end at _200,
  please change the datanames accordingly if you modify the number of bootstraps */

  DATA model&i ;
  SET whi.&&dataname.&i._0_200 (WHERE = (month ne .)) /* whi.&&dataname.&i._101_200 (WHERE = (month ne .)) */ ;
  risk_diff = s0 - s1;
  RUN ;

  PROC TRANSPOSE DATA = model&i OUT = tcond&i PREFIX = riskdiff ;
  VAR risk_diff ;
  ID month ;
  BY bsample ;
  RUN;

  PROC TRANSPOSE DATA = model&i OUT = s0 PREFIX = s0_ ;
  VAR s0 ;
  ID month ;
  BY bsample ;
  RUN;

```

```
PROC TRANSPOSE DATA = model&i OUT = s1 PREFIX = s1_ ;
VAR s1 ;
ID month ;
BY bsample ;
RUN;
```

```
DATA tcond&i;
MERGE tcond&i s0 s1 ;
BY bsample ;
DROP _NAME_ _LABEL_ ;
RUN;
```

```
TITLE "results for cond : &&cond&i ";
```

```
PROC UNIVARIATE DATA = tcond&i (WHERE = (bsample > 0)) ;
VAR riskdiff24 riskdiff72 riskdiff96 ;
RUN;
```

```
PROC PRINT DATA = tcond&i NOOBS ;
RUN;
```

```
%END ;
```

```
%do i = 1 %to 7 ;
```

```
PROC MEANS DATA = tcond&i (WHERE = (bsample > 0)) MEAN STD MIN MAX P5 P95;
VAR riskdiff24 riskdiff72 riskdiff96;
TITLE "results for cond : &&cond&i " ;
OUTPUT out = _conf&i(DROP = _TYPE_ _FREQ_) STD( riskdiff24 riskdiff72 riskdiff96 ) =std24 std72 std96 ;
RUN;
```

```
DATA _cc&i ;
MERGE tcond&i ( KEEP = bsample riskdiff24 riskdiff72 riskdiff96 WHERE = (bsample = 0 )) _conf&i;
LENGTH cond $30 ;
cond = "&&cond&i" ;
ARRAY lb{3} lb24 lb72 lb96 ;
ARRAY ub{3} ub24 ub72 ub96 ;
ARRAY std{3} std24 std72 std96 ;
ARRAY c{3} riskdiff24 riskdiff72 riskdiff96 ;
DO i = 1 TO 3;
lb[i] = ROUND(100*(c[i] - 1.96 * std[i]),0.01);
ub[i] = ROUND(100*(c[i] + 1.96 * std[i]),0.01);
c[i] = ROUND(100*c[i],0.01);
END;
```

```
LABEL riskdiff24 = "RD at 2 years"
riskdiff72 = "RD at 6 years"
riskdiff96 = "RD at 8 years"
lb24 = "Lower 95%"
lb72 = "Lower 95%"
lb96 = "Lower 95%"
ub24 = "Upper 95%"
ub72 = "Upper 95%"
ub96 = "Upper 95%"
cond = "Subset model"
;
```

```
DROP i bsample ;
RUN;
```

```
DATA table4rd ;
SET table4rd _cc&i ;
IF lb24 > . ;
RUN;
```

```
%end;
```

```
TITLE ;
PROC PRINT DATA = table4rd NOOBS LABEL ;
VAR cond riskdiff24 lb24 ub24 riskdiff72 lb72 ub72 riskdiff96 lb96 ub96 ;
RUN;
```

```
%MEND results;
```

```
/*
IDENTIFY LOCATION OF LIBRARY, CALL OUT MACRO FOR BOOTSTRAP ANALYSIS
*/
```

```
LIBNAME whi 'c:\whi';
```

```
options nomprint nonotes;
```

```
%run_all(bootstrap = 1, /* Run bootstrap analysis 0=no, 1=yes */
nboot = 200, /* Number of bootstrap samples to run in total */
bootstart = 0, /* Bootstrap sample to start with 0 = original sample, must be less than nboot */
bootend = 200, /* last bootstrap sample to run */
keep_listing = 0, /* keep output listing for bootstrap samples 0=no, 1=yes */
dataname=linear); /* prefix name for dataset to contain individual bootstrap sample results */
```

```
/******  
  OUTPUT BOOTSTRAP RESULTS  
  *****/  
%let dataname=linear;  
%results;
```

PROGRAM 4  
SURVIVAL CURVES

This program creates the invasive breast cancer-free survival curves for all women in the study.  
The analysis uses the permanent SAS dataset, ipwdr\_bc, created in Program 1 Data management.

```
%MACRO initialize_data;
DATA ipw_a;
SET whi.ipwdr_bc (KEEP=
    id r invasive invady invayear death deathdy dyear lastcont /* outcome related */
    adhrate estrogen progest estpro openlabel pillfreq1 pillfreq2 /* hormone use */
    year dmflag /* year and DM status */
    region ethnic educ marital /* demographic factors */
    menarche gravid parity brstfed brstdisb agefbir /* reproductive factors */
    bcnun brcafrel bkbonrel /* family history */
    age ager diabtrtb smokingb metb bmicxb alcoholb /* established risk factors */
    genhelb bodpainb physfunb preventb hopefulb /* physical & mental health */
    sermb bisphob multib fruitb vegeb /* medication & diet */
    chdb cardiob cancerb kbboneb osteopb /* cvd, cancer, fracture */
    boophb nomam2yr /* mammogram & breast exam */
    menoage menosymb vasomotorb /* menopause-related */
    recency tothcat oc /* hrt and oc use */

    ldiabslf lsmoking lmet lbmicx lalcohol /* established risk factors */
    lgenhel lbodpain lphysfun lprevent /* physical & mental health */
    lserm lbispho lmulti /* medication & diet */
    lchd lcardio lcancer lfracadj losteop /* cvd, cancer, fracture */
    looph lbrstexam lmammogrm lbrstbpsy lbrfin lbrmafin /* mammogram & breast exam */
    lmenosym lvasomotor /* menopause-related */
    lvagbleed lheavybld lintermit lbleednow lbrsttend lsevtend lbrstchg); /* side effects */

/*-----
   RECODE VARIABLES
   -----*/
region2=(region=2);      region3=(region=3);      region4=(region=4);
ethnic3=(ethnic=3);     ethnic4=(ethnic=4);     ethnic8=(ethnic=8);
educ1=(educ=1);        educ2=(educ=2);        educ9=(educ=9);
marital1=(marital=1);  marital2=(marital=2);  marital3=(marital=3);
ageg2=(55<=age<60);    ageg3=(60<=age<65);    ageg4=(65<=age<70);    ageg5=(70<=age<75);
ageg6=(75<=age<80);
smokingb1=(smokingb=1); smokingb2=(smokingb=2); smokingb9=(smokingb=9);
metb1=(metb=1);        metb2=(metb=2);        metb3=(metb=3);        metb4=(metb=4);
metb9=(metb=9);
bmicxb3=(bmicxb=3);    bmicxb4=(bmicxb=4);    bmicxb5=(bmicxb=5);    bmicxb6=(bmicxb=6);
alcoholb2=(alcoholb=2); alcoholb3=(alcoholb=3); alcoholb4=(alcoholb=4); alcoholb5=(alcoholb=5);
alcoholb6=(alcoholb=6); alcoholb9=(alcoholb=9);
genhelb2=(genhelb=2);  genhelb3=(genhelb=3);  genhelb4=(genhelb=4);  genhelb5=(genhelb=5);
bodpainb2=(bodpainb=2); bodpainb3=(bodpainb=3); bodpainb4=(bodpainb=4); bodpainb5=(bodpainb=5);
bodpainb9=(bodpainb=9);
physfunb2=(physfunb=2); physfunb3=(physfunb=3); physfunb4=(physfunb=4);
preventb1=(preventb=1); preventb2=(preventb=2);
hopefulb1=(hopefulb=1); hopefulb2=(hopefulb=2); hopefulb3=(hopefulb=3); hopefulb5=(hopefulb=5);
hopefulb9=(hopefulb=9);
fruitb1=(fruitb=1);    fruitb2=(fruitb=2);    fruitb3=(fruitb=3);
vegeb1=(vegeb=1);      vegeb2=(vegeb=2);      vegeb3=(vegeb=3);
menoage1=(menoage=1);  menoage2=(menoage=2);  menoage9=(menoage=9);
menosymb1=(menosymb=1); menosymb2=(menosymb=2); menosymb3=(menosymb=3);
recency1=(recency=1);  recency2=(recency=2);  recency3=(recency=3);  recency4=(recency=4);
tothcat1=(tothcat=1);  tothcat2=(tothcat=2);  tothcat3=(tothcat>=3);

lsmoking1=(lsmoking=1); lsmoking9=(lsmoking=9);
lmet1=(lmet=1);        lmet2=(lmet=2);        lmet3=(lmet=3);        lmet4=(lmet=4);
lmet9=(lmet=9);
lbmicx1=(lbmicx=1);    lbmicx3=(lbmicx=3);    lbmicx4=(lbmicx=4);    lbmicx5=(lbmicx=5);
lbmicx6=(lbmicx=6);    lbmicx9=(lbmicx=9);
lalcohol2=(lalcohol=2); lalcohol3=(lalcohol=3); lalcohol4=(lalcohol=4); lalcohol5=(lalcohol=5);
lalcohol6=(lalcohol=6); lalcohol9=(lalcohol=9);
lgenhel2=(lgenhel=2);  lgenhel3=(lgenhel=3);  lgenhel4=(lgenhel=4);  lgenhel5=(lgenhel=5);
lgenhel9=(lgenhel=9);
lbodpain2=(lbodpain=2); lbodpain3=(lbodpain=3); lbodpain4=(lbodpain=4); lbodpain5=(lbodpain=5);
lbodpain9=(lbodpain=9);
lphysfun2=(lphysfun=2); lphysfun3=(lphysfun=3); lphysfun4=(lphysfun=4);
lprevent1=(lprevent=1); lprevent2=(lprevent=2); lprevent3=(lprevent=3); lprevent4=(lprevent=4);
lprevent5=(lprevent=5); lprevent6=(lprevent=6);
lcardio1=(lcardio=1);  lcardio2=(lcardio=2);  lcardio3=(lcardio=3);
lmenosym1=(lmenosym=1); lmenosym2=(lmenosym=2); lmenosym3=(lmenosym=3);
lheavybld1=(lheavybld=1); lheavybld2=(lheavybld=2); lheavybld3=(lheavybld=3); lheavybld4=(lheavybld=4);
lsevtend1=(lsevtend=1); lsevtend2=(lsevtend>=2);

menarchel=(menarche=1); menarche2=(menarche=2);
gravid1=(gravid=1);     gravid2=(gravid=2);     gravid3=(gravid=3);     gravid9=(gravid=9);
parity1=(parity=1);     parity2=(parity in (2,3)); parity3=(parity in (4,5)); parity4=(parity>5);
```

```
parity9=(parity=9);
brstfed1=(brstfed=1);      brstfed2=(brstfed=2);
brstdisb1=(brstdisb=1);   brstdisb2=(brstdisb=2);   brstdisb9=(brstdisb=9);
agefbir1=(agefbir=1);     agefbir2=(agefbir=2);     agefbir3=(agefbir=3);     agefbir9=(agefbir=9);
vasomotorb1=(vasomotorb=1);vasomotorb2=(vasomotorb>1);
```

```
lbrfin0=(lbrfin=0);      lbrfin1=(lbrfin=1);      lbrfin2=(lbrfin=2);
lmafin0=(lmafin=0);     lmafin1=(lmafin=1);     lmafin2=(lmafin=2);
lbrmafin0=(lbrmafin=0); lbrmafin1=(lbrmafin=1); lbrmafin2=(lbrmafin=2);
lvasomotor1=(lvasomotor=1);lvasomotor2=(lvasomotor>1);
```

```
DROP smokingb metb bmicxb region ethnic educ marital hopefulb genhelb bodpainb fruitb vegeb preventb menosymb
physfunb alcoholb
lsmoking lmet lbmicx lgenhel lbodpain lcardio lphysfun lalcohol lmenosym lheavybld lsevtend lprevent
menarche parity brstfed brstdisb lbrfin lmafin lbrmafin; RUN;
```

```
/*-----
  ASSIGN ADHRATE FOR THOSE W/ MISSING VALUES
-----*/
```

```
PROC SORT DATA=ipw_a;
BY id year;
RUN;
```

```
DATA ipw_b;
SET ipw_a (RENAME=(adhrate=adhrate_old));
BY id year;
adhrate = adhrate_old;
d      = RANUNI(1232);
```

```
IF adhrate_old=. THEN DO;
  IF pillfreq1=0 THEN DO;
    IF pillfreq2=0 THEN adhrate = (0.0 + 0.0) / 14;      /* not at all          */
    IF pillfreq2=1 THEN adhrate = (0.0 + 0.5) / 14;      /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (0.0 + 1.5) / 14;      /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (0.0 + 3.5) / 14;      /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (0.0 + 5.5) / 14;      /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (0.0 + 7.0) / 14;      /* every day of the week */
    IF pillfreq2=. THEN adhrate = (0.0 + 7*d) / 14;      /* missing              */
  END;
  IF pillfreq1=1 THEN DO;
    IF pillfreq2=0 THEN adhrate = (0.5 + 0.0) / 14;      /* not at all          */
    IF pillfreq2=1 THEN adhrate = (0.5 + 0.5) / 14;      /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (0.5 + 1.5) / 14;      /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (0.5 + 3.5) / 14;      /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (0.5 + 5.5) / 14;      /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (0.5 + 7.0) / 14;      /* every day of the week */
    IF pillfreq2=. THEN adhrate = (0.5 + 7*d) / 14;      /* missing              */
  END;
  IF pillfreq1=2 THEN DO;
    IF pillfreq2=0 THEN adhrate = (1.5 + 0.0) / 14;      /* not at all          */
    IF pillfreq2=1 THEN adhrate = (1.5 + 0.5) / 14;      /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (1.5 + 1.5) / 14;      /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (1.5 + 3.5) / 14;      /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (1.5 + 5.5) / 14;      /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (1.5 + 7.0) / 14;      /* every day of the week */
    IF pillfreq2=. THEN adhrate = (1.5 + 7*d) / 14;      /* missing              */
  END;
  IF pillfreq1=3 THEN DO;
    IF pillfreq2=0 THEN adhrate = (3.5 + 0.0) / 14;      /* not at all          */
    IF pillfreq2=1 THEN adhrate = (3.5 + 0.5) / 14;      /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (3.5 + 1.5) / 14;      /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (3.5 + 3.5) / 14;      /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (3.5 + 5.5) / 14;      /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (3.5 + 7.0) / 14;      /* every day of the week */
    IF pillfreq2=. THEN adhrate = (3.5 + 7*d) / 14;      /* missing              */
  END;
  IF pillfreq1=4 THEN DO;
    IF pillfreq2=0 THEN adhrate = (5.5 + 0.0) / 14;      /* not at all          */
    IF pillfreq2=1 THEN adhrate = (5.5 + 0.5) / 14;      /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (5.5 + 1.5) / 14;      /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (5.5 + 3.5) / 14;      /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (5.5 + 5.5) / 14;      /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (5.5 + 7.0) / 14;      /* every day of the week */
    IF pillfreq2=. THEN adhrate = (5.5 + 7*d) / 14;      /* missing              */
  END;
  IF pillfreq1=5 THEN DO;
    IF pillfreq2=0 THEN adhrate = (7.0 + 0.0) / 14;      /* not at all          */
    IF pillfreq2=1 THEN adhrate = (7.0 + 0.5) / 14;      /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (7.0 + 1.5) / 14;      /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (7.0 + 3.5) / 14;      /* 3-4 days per week   */
    IF pillfreq2=4 THEN adhrate = (7.0 + 5.5) / 14;      /* 5-6 days per week   */
    IF pillfreq2=5 THEN adhrate = (7.0 + 7.0) / 14;      /* every day of the week */
    IF pillfreq2=. THEN adhrate = (7.0 + 7*d) / 14;      /* missing              */
  END;
  IF pillfreq1=. THEN DO;
    IF pillfreq2=0 THEN adhrate = (7*d + 0.0) / 14;      /* not at all          */
    IF pillfreq2=1 THEN adhrate = (7*d + 0.5) / 14;      /* < 1 day per week    */
    IF pillfreq2=2 THEN adhrate = (7*d + 1.5) / 14;      /* 1-2 days per week   */
    IF pillfreq2=3 THEN adhrate = (7*d + 3.5) / 14;      /* 3-4 days per week   */

```

```

IF pillfreq2=4 THEN adhrate = (7*d + 5.5) / 14; /* 5-6 days per week */
IF pillfreq2=5 THEN adhrate = (7*d + 7.0) / 14; /* every day of the week */
IF pillfreq2=. THEN adhrate = d;
END;
END;

IF 0<adhrate<=0.01 THEN adhrate=0; /* if taking <1% in a given year, then recode as taking none */
IF adhrate> 1 THEN adhrate=1; /* if taking >100% in a given year, then recode as taking one */

/*-----
ASSIGN DOSE (INCLUDING NON-STUDY HORMONE USE
-----*/
extra=0;
dose_e=0;
IF (estrogen=1 & progest=1) OR estpro=1 OR openlabel=1 THEN DO;
extra=1; dose_e=RANUNI(2321);
END;

IF r=1 THEN DO;
IF extra=0 THEN dose=adhrate;
IF extra=1 THEN dose=adhrate + dose_e;
END;

IF r=0 THEN DO;
IF extra=0 THEN dose=0;
IF extra=1 THEN dose=dose_e;
END;

/*-----
CREATE VARIABLES FOR WEIGHT ESTIMATION
-----*/
IF dose=0 THEN cendose=0;
ELSE cendose=1; /* cendose: censoring indicator for those w/ dose=0 */
IF dose>0 THEN logdose=LOG(dose); /* logdose: log-transformed dose for those w/ dose>0 */
ldose = LAG(dose);
IF FIRST.id THEN ldose=dose; /* hormone dose in the previous year */

DROP adhrate_old d; RUN;

%MEND initialize_data;

/*****
WEIGHT ESTIMATION
*****/

%MACRO estimate_weights;

/*-----
PROBABILITY OF HAVING DOSE=0
-----*/
/* model 1a - numerator of sw */
PROC LOGISTIC DATA=ipw_b (WHERE= ( R=0 ));
MODEL cendose=year year*year
dmflag ageg2--ageg6 ethnic3--ethnic8
bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
oc brcafrel
tothcat1--tothcat3
menarchel--menarche2
marital1--marital3 menoage1--menoage9
ldose ldose*ldose ldose*ldose*ldose;
OUTPUT OUT=modella1 P=num_adh;
RUN;

PROC LOGISTIC DATA=ipw_b(WHERE = ( R=1)) ;
MODEL cendose=year year*year
dmflag ageg2--ageg6 ethnic3--ethnic8
bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
oc brcafrel
tothcat1--tothcat3
menarchel--menarche2
marital1--marital3 menoage1--menoage9
ldose ldose*ldose ldose*ldose*ldose;
OUTPUT OUT=modella2 P=num_adh;
RUN;

/* model 1b - denominator of sw */
PROC LOGISTIC DATA=ipw_b(WHERE=( R=0)) ;
MODEL cendose=year year*year
dmflag ageg2--ageg6 ethnic3--ethnic8
bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
oc brcafrel bkbonrel
educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
menarchel--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9

```

```
physfunb2--physfunb4 preventb1--preventb2
chdb cardiob cancerb bkboneb osteopb
```

```
lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
lchd lcardio1--lcardio3 lcancer lfracadj lsteop
lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
lbrsttend lsevtend1--lsevtend2 lbrstchn9
ldose ldose*ldose ldose*ldose*ldose;
```

```
OUTPUT OUT=model1b1 P=den_adh;
RUN;
```

```
PROC LOGISTIC DATA=ipw_b(WHERE= (R=1));
```

```
MODEL cendose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
  oc brcafrel bkbonrel
  educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
  menarchel1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
  region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
  physfunb2--physfunb4 preventb1--preventb2
  chdb cardiob cancerb bkboneb osteopb
```

```
lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
lchd lcardio1--lcardio3 lcancer lfracadj lsteop
lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
lbrsttend lsevtend1--lsevtend2 lbrstchn9
ldose ldose*ldose ldose*ldose*ldose;
```

```
OUTPUT OUT=model1b2 P=den_adh;
RUN;
```

```
/*-----
  PROBABILITY OF HAVING A PARTICULAR DOSE AMONG THOSE W/ DOSE>0
-----*/
```

```
/* model 2a - numerator of sw */
```

```
PROC GLM DATA=ipw_b(WHERE=(cendose=1 AND R=0)) ORDER=FREQ
  OUTSTAT=sse2a1(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
```

```
MODEL logdose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafrel
  tothcat1--tothcat3
  menarchel1--menarche2
  marital1--marital3 menoage1--menoage9
  ldose ldose*ldose ldose*ldose*ldose / solution;
```

```
OUTPUT OUT=model2a1_1 P=pred;
RUN;
QUIT;
```

```
PROC GLM DATA=ipw_b(WHERE=( cendose=1 AND R=1)) ORDER=FREQ
```

```
  OUTSTAT=sse2a2(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
```

```
MODEL logdose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
  oc brcafrel
  tothcat1--tothcat3
  menarchel1--menarche2
  marital1--marital3 menoage1--menoage9
  ldose ldose*ldose ldose*ldose*ldose / solution;
```

```
OUTPUT OUT=model2a2_1 P=pred;
RUN;
QUIT;
```

```
/* MODEL 2b - denominator of sw */
```

```
PROC GLM DATA=ipw_b(WHERE=( cendose=1 AND R=0)) ORDER=FREQ
```

```
  OUTSTAT=sse2b1(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
```

```
MODEL logdose=year year*year
  dmflag ageg2--ageg6 ethnic3--ethnic8
  bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
  oc brcafrel bkbonrel
  educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
  menarchel1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
  region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
  physfunb2--physfunb4 preventb1--preventb2
  chdb cardiob cancerb bkboneb osteopb
```

```
lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
lchd lcardio1--lcardio3 lcancer lfracadj lsteop
```



```

lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
lbrsttend lsevtend1--lsevtend2 lbrstchng
ldose ldose*ldose ldose*ldose*ldose / solution;
OUTPUT OUT=model2b1_1 P=pred R=residual;
RUN;
QUIT;

PROC GLM DATA=ipw_b(WHERE =(cendose=1 AND R=1)) ORDER=FREQ
  OUTSTAT=sse2b2(KEEP=_SOURCE_ _TYPE_ DF SS WHERE=( _SOURCE_ IN('ERROR') AND _TYPE_ IN('ERROR')));
MODEL logdose=year year*year
  dmflag age2--age6 ethnic3--ethnic8
  bmicx3--bmicx6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
  parity1--parity9 agefbir1--agefbir3 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
  oc brcafrel bkbbonrel
  educ1--educ9 genhelb2--genhelb5 boophb tothcat1--tothcat3
  menarchel1--menarche2 brstdisb1--brstdisb9 brstfed1--brstfed2
  region2--region4 marital1--marital3 fruitb1--fruitb3 vegeb1--vegeb3 menoage1--menoage9
  physfunb2--physfunb4 preventb1--preventb2
  chdb cardiob cancerb bkboneb osteopb

  lbmicx1--lbmicx9 lmet1--lmet9 lsmoking1 lsmoking9 lalcohol2--lalcohol9
  looph lmenosym1--lmenosym3 lvasomotor1--lvasomotor2
  lbrstexam lmammogrm lbrstbpsy lbrmafin0--lbrmafin2
  lgenhel2--lgenhel9 lphysfun2--lphysfun4 lprevent1--lprevent6
  lchd lcardio1--lcardio3 lcancer lfracadj lsteop
  lvagbleed lheavybld1--lheavybld3 lintermit lbleednow
  lbrsttend lsevtend1--lsevtend2 lbrstchng
  ldose ldose*ldose ldose*ldose*ldose / solution;
OUTPUT OUT=model2b2_1 P=pred R=residual;
RUN;
QUIT;

DATA sse2a1;
SET sse2a1;
rootmse_n=SQRT(ss/df);
RUN;

DATA sse2a2;
SET sse2a2;
rootmse_n=SQRT(ss/df);
RUN;

DATA sse2b1;
SET sse2b1;
rootmse_n=SQRT(ss/df);
RUN;

DATA sse2b2;
SET sse2b2;
rootmse_n=SQRT(ss/df);
RUN;

PROC SQL;
  CREATE TABLE model2a1 AS
  SELECT O.*, L.rootmse_n
  FROM model2a1_1 O, sse2a1 L;
QUIT;

PROC SQL;
  CREATE TABLE model2a2 AS
  SELECT O.*, L.rootmse_n
  FROM model2a2_1 O, sse2a2 L;
QUIT;

PROC SQL;
  CREATE TABLE model2b1 AS
  SELECT O.*, L.rootmse_n
  FROM model2b1_1 O, sse2b1 L;
QUIT;

PROC SQL;
  CREATE TABLE model2b2 AS
  SELECT O.*, L.rootmse_n
  FROM model2b2_1 O, sse2b2 L;
QUIT;

DATA model2a1;
SET model2a1;
num_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);
FORMAT num_adhrate 12.10;
RUN;

DATA model2a2;
SET model2a2;
num_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);
FORMAT num_adhrate 12.10;
RUN;

DATA model2b1;
SET model2b1;
den_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);

```

```
FORMAT den_adhrate 12.10;
RUN;
```

```
DATA model2b2;
SET model2b2;
den_adhrate=PDF('NORMAL',logdose,pred,rootmse_n);
FORMAT den_adhrate 12.10;
RUN;
```

```
/*-----
  ESTIMATE STABILIZED WEIGHT
-----*/
```

```
DATA ipw1;
SET ipw_b (KEEP=id r adhrate dose dose_e extra cendose year invasive invayear invady death dyear deathdy lastcont
          ager menoaage dmflag age age2--age6 ethnic3--ethnic8 bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9
          alcoholb2--alcoholb9 parity1--parity9 nomam2yr menosymb1--menosymb3 vasomotorb1--vasomotorb2
          oc brcafrel tothcat tothcat1--tothcat3 menarchel1--menarche2 marital1--marital3 menoaage1--menoaage9);
RUN;
```

```
DATA modella1;
SET modella1 (KEEP=id year num_adh);
RUN;
```

```
DATA modella2;
SET modella2 (KEEP=id year num_adh);
RUN;
```

```
DATA modellb1;
SET modellb1 (KEEP=id year den_adh);
RUN;
```

```
DATA modellb2;
SET modellb2 (KEEP=id year den_adh);
RUN;
```

```
DATA model2a1;
SET model2a1 (KEEP=id year num_adhrate);
RUN;
```

```
DATA model2a2;
SET model2a2 (KEEP=id year num_adhrate);
RUN;
```

```
DATA model2b1;
SET model2b1 (KEEP=id year den_adhrate);
RUN;
```

```
DATA model2b2;
SET model2b2 (KEEP=id year den_adhrate);
RUN;
```

```
PROC SORT DATA=ipw1;
BY id year;
RUN;
```

```
PROC SORT DATA=modella1;
BY id year;
RUN;
```

```
PROC SORT DATA=modella2;
BY id year;
RUN;
```

```
PROC SORT DATA=modellb1;
BY id year;
RUN;
```

```
PROC SORT DATA=modellb2;
BY id year;
RUN;
```

```
PROC SORT DATA=model2a1;
BY id year;
RUN;
```

```
PROC SORT DATA=model2a2;
BY id year;
RUN;
```

```
PROC SORT DATA=model2b1;
BY id year;
RUN;
```

```
PROC SORT DATA=model2b2;
BY id year;
RUN;
```

```
DATA ipw;
MERGE ipw1 modella1 modella2 modellb1 modellb2 model2a1 model2a2 model2b1 model2b2;
BY id year;
IF FIRST.id THEN stabw=1;
RETAIN stabw;
```

```
w2 = num_adh / den_adh;
w3 = (1 - num_adh) / (1 - den_adh);
w4 = num_adhrate / den_adhrate;
```

```
IF cendose=0 THEN stabw = stabw * w2;
IF cendose=1 THEN stabw = stabw * w3 * w4;
```

```
DROP w2-w4;
RUN;
```

```
DATA weight;
SET ipw (KEEP=year stabw);
```

```

RUN;

PROC SORT DATA=weight;
BY year;
RUN;

TITLE2 'dist of weights by year';
PROC MEANS MAXDEC=4 DATA=weight N MEAN STD MIN P1 P25 P50 P75 P99 MAX;
VAR stabw;
BY year;
RUN;

TITLE2 'dist of weights';
PROC MEANS MAXDEC=4 DATA=weight N MEAN STD MIN P1 P25 P50 P75 P99 MAX;
VAR stabw;
RUN;

PROC DATASETS LIBRARY=work NOLIST;
DELETE modella1 modella2 modellb1 modellb2 model2a1 model2a2 model2b1 model2b2
model2a1_1 model2a2_1 model2b1_1 model2b2_1 sse2a1 sse2a2 sse2b1 sse2b2 ;
QUIT;

%MEND estimate_weights;

/*****
EXPAND DATASET INTO PERSON-MONTH STRUCTURE
*****/

%MACRO create_final_dataset;

/* create 'maxyear' to be used for do loop */
PROC SQL;
CREATE TABLE a AS
SELECT id, MAX(year) AS maxyear
FROM ipw GROUP BY id;
QUIT;

PROC SQL;
CREATE TABLE b AS
SELECT O.*, L.maxyear
FROM ipw O, a L WHERE O.id=L.id;
QUIT;

/* create month of invasive breast cancer (invamo) */
DATA b;
SET b;
invamo = CEIL(invady/30.4375);
deathmo = CEIL(deathdy/30.4375);

mm=0; RETAIN mm;
DO y=1 UNTIL (year=maxyear);
  DO m=1 TO 12;
    mm=mm+1;
    OUTPUT;
  END;
END; DROP y m; RUN;

/* create months since randomization (month) & delete observation after bc or death occurred, or last follow-up month */
DATA b; SET b;
month=mm+12*(year-1);

maxmo = maxyear*12;
lmo = CEIL(lastcont/30.4375);
lyear = CEIL(lastcont/365.25);
lastmo = MIN(lmo,maxmo); /* lmo: last month based on lastcont; maxmo: last month based on data */
lastyear = MIN(lyear,maxyear); /* lyear: last year based on lastcont; maxyear: last year based on data */

IF invamo NE . AND month > invamo THEN DELETE;
IF deathmo NE . AND month > deathmo THEN DELETE;
IF invamo = . AND deathmo = . AND month > lastmo THEN DELETE;

IF invamo NE . AND invamo = month THEN invasive=1;
ELSE invasive=0;
IF deathmo NE . AND deathmo = month THEN death=1;
ELSE death=0;

DROP mm;
RUN;

/*-----
CALCULATE CUMULATIVE USE
-----*/

DATA p;
SET whi.baseline_bc(KEEP=id);
p=RANUNI(99999);
RUN;

```

```

DATA b;
MERGE p b;
BY id;
RUN;

PROC SORT DATA=b;
BY id month;
RUN;

DATA b;
SET b;
BY id month;

IF FIRST.id THEN cuma = 0;
RETAIN cuma;

IF (year NE lastyear) & (year NE invayear) & (year NE dyear) THEN DO;
  cuma = cuma + (dose / 12);
END;

IF (year=lastyear) & (year NE invayear) & (year NE dyear) THEN DO;
  d0 = (lastmo-((lastyear-1)*12));
  cuma = cuma + (dose / d0);
END;

IF year = dyear THEN DO;
  d1 = (deathmo-((dyear-1)*12));
  cuma = cuma + (dose / d1);
END;

IF (year = invayear) & (invayear NE lastyear) & (year NE dyear) THEN DO;
  d2 = invamo-((invayear-1)*12);
  d3 = 13 - d2;
  IF 0/d3 <= p <= 1/d3 THEN cuma = cuma + (dose / d2);
  IF 1/d3 < p <= 2/d3 THEN cuma = cuma + (dose / (d2+1));
  IF 2/d3 < p <= 3/d3 THEN cuma = cuma + (dose / (d2+2));
  IF 3/d3 < p <= 4/d3 THEN cuma = cuma + (dose / (d2+3));
  IF 4/d3 < p <= 5/d3 THEN cuma = cuma + (dose / (d2+4));
  IF 5/d3 < p <= 6/d3 THEN cuma = cuma + (dose / (d2+5));
  IF 6/d3 < p <= 7/d3 THEN cuma = cuma + (dose / (d2+6));
  IF 7/d3 < p <= 8/d3 THEN cuma = cuma + (dose / (d2+7));
  IF 8/d3 < p <= 9/d3 THEN cuma = cuma + (dose / (d2+8));
  IF 9/d3 < p <= 10/d3 THEN cuma = cuma + (dose / (d2+9));
  IF 10/d3 < p <= 11/d3 THEN cuma = cuma + (dose / (d2+10));
  IF 11/d3 < p <= 12/d3 THEN cuma = cuma + (dose / (d2+11));
END;

IF (year = invayear) & (invayear=lastyear) & (year NE dyear) THEN DO;
  d4 = invamo-((invayear-1)*12);
  d5 = lastmo-((lastyear-1)*12);
  d6 = (d5-d4) + 1;

  IF 0/d6 <= p <= 1/d6 THEN cuma = cuma + (dose / d4);
  IF 1/d6 < p <= 2/d6 THEN cuma = cuma + (dose / (d4+1));
  IF 2/d6 < p <= 3/d6 THEN cuma = cuma + (dose / (d4+2));
  IF 3/d6 < p <= 4/d6 THEN cuma = cuma + (dose / (d4+3));
  IF 4/d6 < p <= 5/d6 THEN cuma = cuma + (dose / (d4+4));
  IF 5/d6 < p <= 6/d6 THEN cuma = cuma + (dose / (d4+5));
  IF 6/d6 < p <= 7/d6 THEN cuma = cuma + (dose / (d4+6));
  IF 7/d6 < p <= 8/d6 THEN cuma = cuma + (dose / (d4+7));
  IF 8/d6 < p <= 9/d6 THEN cuma = cuma + (dose / (d4+8));
  IF 9/d6 < p <= 10/d6 THEN cuma = cuma + (dose / (d4+9));
  IF 10/d6 < p <= 11/d6 THEN cuma = cuma + (dose / (d4+10));
  IF 11/d6 < p <= 12/d6 THEN cuma = cuma + (dose / (d4+11));
END;
RUN;

/*****
CREATE PERMANENT DATASET FOR ANALYSIS
*****/
DATA whi.dr_bc;
SET b (KEEP= cuma stabw id r month year invasive ager menoaage dmflag ageg2--ageg6 ethnic3--ethnic8
      bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9 parity1--parity9
      nomam2yr vasomotorb1--vasomotorb2 oc brcafre1 tothcat tothcat1--tothcat3 menarche1--menarche2
      marital1--marital3 menoaage1--menoaage9);
RUN;

PROC DATASETS LIBRARY=work NOLIST ;
DELETE a b p ;
QUIT;

%MEND create_final_dataset;

/*-----
MACRO FOR CUBIC SPLINES
-----*/
%macro rcspline(x,knot1,knot2,knot3,knot4,knot5,knot6,knot7,knot8,knot9,knot10, norm=2) ;

```

```

%LOCAL j v7 k tk tk1 t k1 k2;
%LET v7=&x; %IF %length(&v7)=8 %THEN %LET v7=%SUBSTR(&v7,1,7);
%*get no. knots, last knot, next to last knot;
%DO k=1 %TO 10;
%IF %QUOTE(&&knot&k)= %THEN %GOTO nomorek;
%END;
%LET k=11;
%nomorek: %LET k=%EVAL(&k-1); %LET k1=%EVAL(&k-1); %LET k2=%EVAL(&k-2);
%IF &k<3 %THEN %PUT ERROR: <3 knots given. no spline variables CREATED.;
%ELSE %DO;
%LET tk=&&knot&k;
%LET tk1=&&knot&k1;
DROP _kd_ _kd_=
%IF &norm=0 %THEN 1;
%ELSE %IF &norm=1 %THEN &tk - &tk1;
%ELSE (&tk - &knot1)**.6666666666666666; ;
%DO j=1 %TO &k2;
%LET t=&&knot&j;
&v7&j=max((&x-&t)/_kd_,0)**3+((&tk1-&t)*max((&x-&tk)/_kd_,0)**3
-(&tk-&t)*max((&x-&tk1)/_kd_,0)**3)/(&tk-&tk1)%STR(;);
%END;
%END;
%mend rcspline;

```

```
LIBNAME whi 'c:\whi';
```

```

%initialize_data;
%estimate_weights;
%create_final_dataset;

```

```

DATA r; SET whi.dr_bc;
%rcspline(month,6,48,90);
interact1 = cuma * month;
interact2 = cuma * month1;

```

```

a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;
RUN;

```

```

PROC LOGISTIC DESCENDING DATA=r;
ODS OUTPUT ParameterEstimates=pe;
MODEL invasive = cuma
month month1
dmflag ageg2--ageg6 ethnic3--ethnic8
bmicxb3--bmicxb6 metb1--metb9 smokingb1--smokingb9 alcoholb2--alcoholb9
parity1--parity9 nomam2yr vasomotorb1--vasomotorb2
oc brcafre1
tothcat1--tothcat3
menarchel1--menarche2
marital1--marital3 menoage1--menoage9
interact1-interact2 a b;

```

```
WEIGHT stabw;
RUN;
```

```

PROC SQL NOPRINT; SELECT estimate FORMAT=16.12 INTO:ibc_estimate SEPARATED BY ' ' FROM pe; QUIT;
PROC SQL NOPRINT; SELECT variable INTO:model SEPARATED BY ' ' FROM pe; QUIT;

```

```

PROC MEANS SUM NOPRINT DATA=pe; VAR df; OUTPUT OUT=nobs(DROP=_TYPE_ _FREQ_ WHERE=( _STAT_="N")); RUN;
PROC SQL NOPRINT; SELECT df INTO:nvar SEPARATED BY ' ' FROM nobs; QUIT;

```

```
PROC SORT DATA=r; BY id month; RUN;
```

```

DATA r1 (KEEP=month s0 s1); SET r; where month=1;
ARRAY var{&nvar} &model;
ARRAY coef{&nvar} (&ibc_estimate);

```

```

intercept = 1;
s0 = 1;
s1 = 1;

```

```

DO month=1 TO 96;
%rcspline(month,6,48,90);
xbeta0 = 0;
xbetal = 0;
a = (tothcat NE 0) * month;
b = (tothcat NE 0) * month1;

```

```
DO i=1 TO DIM(var);
```

```

/* never treated */
cuma = 0;
interact1 = 0;
interact2 = 0;
xbeta0 = xbeta0 + coef[i] * var[i];

```

```

/* always treated */
cuma      = (month/12);
interact1 = (month/12) * month;
interact2 = (month/12) * month1;
xbetal    = xbetal + coef[i] * var[i];
END;

p0 = 1 / (1 + EXP(-xbeta0));          /* risk of IBC for never treated */
p1 = 1 / (1 + EXP(-xbetal));          /* risk of IBC for always treated */
s0 = s0 * (1 - p0);                   /* IBC-free survival for never treated */
s1 = s1 * (1 - p1);                   /* IBC-free survival for always treated */
OUTPUT;
END; RUN;

PROC MEANS DATA=r1 MEAN NOPRINT; WHERE month NE .;
VAR s0 s1;
OUTPUT OUT=mean_for_curve1(DROP=_TYPE_ _FREQ_) MEAN(s0 s1)=;
CLASS month; RUN;

DATA mean_for_curve1; SET mean_for_curve1;
years      = month / 12;
risk_diff  = (S0 - S1)*100;
LABEL years = 'Years since randomization'
month       = 'Months since randomization'
s0          = 'No use'
s1          = 'Continuous use'
risk_diff   = 'Risk difference';
RUN;

DATA k; month=0; years=0; s0=1; s1=1; risk_diff=0; RUN;
DATA mean_for_curve; MERGE mean_for_curve1 k; BY month; RUN;
DATA mean_for_curve; SET mean_for_curve; WHERE month NE .; RUN;
PROC PRINT DATA=mean_for_curve NOOBS; VAR month s0 s1 risk_diff; RUN;

GOPTIONS RESET=all
          ROTATE=landscape
          CBACK=white
          FTITLE=swissb
          FTEXT=swiss
          HTITLE=4
          HTEXT=3;

GOPTIONS GSFNAME=graphs
          DEVICE=pdfc
          GSFMODE=replace;

AXIS1    ORDER=0.95 TO 1 BY 0.01
          LABEL=(A=90 HEIGHT=3.2 FONT=swiss "Breast cancer-free survival");

SYMBOL1  COLOR=black INTERPOL=stepcj LINE=2 WIDTH=6 VALUE=none;
SYMBOL2  COLOR=black INTERPOL=stepcj LINE=1 WIDTH=10 VALUE=none;

LEGEND1  LABEL=none
          SHAPE=symbol(8,1)
          POSITION=(bottom center inside)
          MODE=share;
FILENAME GRAPHS 'c:\whi\all.pdf';

PROC GPLOT DATA=mean_for_curve;
PLOT s0*years s1*years / OVERLAY HAXIS=0 TO 8 BY 1 HMINOR=0 VAXIS=axis1 VMINOR=0 LEGEND=legend1 NOFRAME; RUN; QUIT;

```